Large sample properties of the three-step euclidean likelihood estimators under model misspecification

Dovonon, Prosper

Concordia University, CIREQ

18 November 2008

Large Sample Properties of the Three-Step Euclidean Likelihood Estimators under Model Misspecification

Prosper DOVONON*
Concordia University and CIREQ

First Draft: November 18, 2008

This Draft: May 16, 2010

Abstract

This paper studies the three-step Euclidean likelihood (3S) estimator and its corrected version as proposed by Antoine, Bonnal and Renault (2007) in globally misspecified models. We establish that the 3S estimator stays $\sqrt{n}$-convergent and asymptotically Gaussian. The discontinuity in the shrinkage factor makes the analysis of the corrected-3S estimator harder to carry out in misspecified models. We propose a slight modification to this factor to control its rate of divergence in case of misspecification. We show that the resulting modified-3S estimator is also higher order equivalent to the maximum empirical likelihood (EL) estimator in well specified models and $\sqrt{n}$-convergent and asymptotically Gaussian in misspecified models. Its asymptotic distribution robust to misspecification is also provided. Because of these properties, both the 3S and the modified-3S estimators could be considered as computationally attractive alternatives to the exponentially tilted empirical likelihood estimator proposed by Schennach (2007) which also is higher order equivalent to EL in well specified models and $\sqrt{n}$-convergent in misspecified models.

Keywords: Misspecified models, Empirical likelihood, Three-step Euclidean likelihood.

*I thank Yuichi Kitamura, and two anonymous referees for very helpful suggestions that led to a much improved version of the article. I thank Bryan Campbell, Marine Carrasco, Sílvia Gonçalves, Alastair Hall, Atsushi Inoue, William McCausland, Nour Meddahi, Benoit Perron, Éric Renault, Enrique Sentana, and Paolo Zaffaroni as well as the seminar participants at the University of Manchester for very useful comments and suggestions, and Patrik Guggenberger for kindly providing me with his computer programs. Address: Department of Economics, Concordia University, 1455 de Maisonneuve Blvd. West, Montreal, Quebec, H3G 1M8 Canada; tel: (514) 848-2424 (ext. 3479), fax: (514) 848-4536, email: prosper.dovonon@concordia.ca.

1 Introduction

The lackluster performance of the generalized method of moments (GMM) estimator in finite samples has paved the way for several competing alternative efficient estimators. Among them, maybe the most known are the continuously updated GMM (CU) estimator proposed by Hansen, Heaton and Yaron (1996) also known to be identical to the Euclidean empirical likelihood (EEL) estimator, the maximum empirical likelihood (EL) estimator proposed by Qin and Lawless (1994) and the exponential tilting (ET) estimator introduced by Kitamura and Stutzer (1997). These estimators are included in both the minimum discrepancy (MD) class of estimators formulated by Corcoran (1998) and the generalized empirical likelihood (GEL) class of estimators proposed by Newey and Smith (2004). When comes to the comparison of these estimators, three points are considered as major issues. The implementation cost, the finite sample bias and the behaviour under model misspecification. All these alternative estimators are computationally very demanding. They are expressed as solutions of saddle point problems and impose a double optimization program solving in their calculation process. (See Kitamura (2006).) When a large parameter vector is considered, these saddle point problems are computationally cumbersome. On the other hand, because of their one-step nature, these estimators have a fewer sources of higher order ($O(n^{-1})$) bias than the efficient two-step GMM and, as shown by Newey and Smith (2004), the EL estimator has even fewer higher order bias sources than all of the other estimators. Still, this enjoyable property of the EL estimator holds only in correctly specified models. A moment condition model is globally misspecified if the true data generating process deviates from these moment conditions such that no value in the parameter space solves the population moment conditions. In the case of global misspecification, Schennach (2007) establishes that EL ceases to be $\sqrt{n}$-convergent whereas ET is $\sqrt{n}$-convergent (Imbens (1997)) and CU may also be $\sqrt{n}$-convergent.

As one can notice, none of these regular estimators enjoys all of the desirable properties. The most recent estimators proposed in the literature aim to combine several of them to deliver new ones with better performance. Schennach (2007) combines the EL and ET estimators to propose the exponentially tilted empirical likelihood (ETEL) estimator which is, in well specified models, higher order equivalent to EL in the sense that their difference is of order $O_p(n^{-3/2})$ and is $\sqrt{n}$-convergent in misspecified models. Still, the ETEL estimator is as computationally costly as both EL and ET. Antoine, Bonnal and Renault (2007) combine the two-step efficient GMM estimator with the Euclidean empirical likelihood implied probabilities to deliver the so-called three-step Euclidean likelihood (3S) estimator. As these implied probabilities could be negative causing some instability to the resulting estimator, they also propose a variant of this estimator which uses some Euclidean likelihood implied probabilities corrected by shrinkage. As highlighted in this paper through our Monte Carlo simulations, such corrections to the implied probabilities are crucial to take any benefit from the three-step procedure. In particular, the 3S estimator appears to be too sensitive to negative implied probabilities and is computationally inefficient with too many outliers in such cases. Nevertheless, these two estimators share two important advantages. They are computationally convenient and are also higher order equivalent to the EL estimator in well specified models.

This paper studies the three-step Euclidean likelihood estimators under global misspecification. Inference under misspecification is getting more and more attention in the econometric literature. White (1982) studies the quasi maximum likelihood estimator when the distributional assumptions are misspecified. Hall (2000) examines the implications of model misspecification for the heteroskedasticity and autocorrelation consistent covariance matrix estimator and the GMM overidentifying restrictions test. Hall and Inoue (2003) study the GMM estimators under global misspecification. Schennach (2007) analyzes the EL and ETEL estimators under global misspecification while Kan and Robotti (2008) propose a methodology to evaluate the Hansen-Jagannathan distance between two pricing kernels in the case of model misspecification.

One of the main motivations for studying estimators in globally misspecified models is underlined by Schennach (2007). Statistical models are only simplification of a complex reality and therefore are bound to be misspecified. Specification tests aim to indicate the candidate models which seem closer to the sample under studies and may be sharp to detect globally misspecified models. Nevertheless, it is frequent to come across parsimonious models delivering better forecasting performances but failing the specification tests while other less parsimonious models pass these tests with very poor out-of-sample performance. For such parsimonious models, $\sqrt{n}$-convergent estimators are useful to allow asymptotic approximations through usual sample sizes. Furthermore, the asymptotic behaviour of these estimators also need to be fully derived.

In the context of moment condition-based models in particular, the specification tests for overidentifying restrictions could validate the model. In the case of rejection, if no theory is available for inference, empirical researchers could have to drop parsimonious, robust and competitive models for forecasting for other less attractive. The situation could even be more ambiguous. Hall and Inoue (2003) report several empirical researches in the literature in which inference by the usual asymptotic distributions have been performed even though the data have rejected the overidentifying restrictions. In this paper, we provide global misspecification robust inference for the 3S estimator. We show that, in the case of moment misspecification, this estimator stays $\sqrt{n}$-convergent and is asymptotically normally distributed and we derive its asymptotic distribution robust to global misspecification.

The main intuition behind this asymptotic behaviour of the 3S estimator under misspecification is related to the fact that its estimating function is equivalent to a smooth function of sample means. This is not the case for the corrected 3S estimator. The discontinuity in the shrinkage factor makes its analysis more difficult. We establish that, under mild conditions, the shrinkage factor diverges such that the estimating function can be considered as an approximation of a smooth function of sample means. But, the asymptotic distribution derivation requires to control the rate of divergence of this factor. For that purpose, we propose a slight modification of the original shrinkage factor proposed by Antoine, Bonnal and Renault (2007) which simplifies the derivations. We call the estimator resulting from the new shrinkage factor the *modified* three-step Euclidean likelihood (m3S) estimator. The m3S estimator is as easy to compute as the 3S estimator. Additionally, we show that in correctly specified models the m3S estimator is higher order equivalent to both the EL and the 3S estimators. As the m3S estimator is computed via implied probabilities corrected for the sign, it is more stable than the 3S estimator. We show that under global misspecification the m3S estimator stays $\sqrt{n}$-convergent and asymptotically Gaussian. Its asymptotic distribution robust to global misspecification is also provided. This makes both the 3S and the m3S estimators two computationally appealing alternative to the ETEL estimator.

The remainder of the paper is organized as follows. Section 2 describes the model and the estimators and establishes the higher order equivalence of the m3S and the EL estimators in well specified models. In Section 3 we derive asymptotic results for the 3S and m3S estimators under moment misspecification. Our Monte Carlo experiments are introduced in Section 4 followed by Section 5 which concludes. All proofs are gathered in the Appendix.

2 The three-step Euclidean likelihood estimators

The statistical model that we consider in this paper is one with finite number of moment restrictions. To describe it, let $\{x_i : i = 1, \ldots, n\}$ be independent realizations of a random vector x and $\psi(x, \theta)$ a known q-vector of functions of the data observation x and the parameter θ which may lie in a compact parameter set $\Theta \subset \mathbb{R}^p$ $(q \geq p)$. We assume in this section that the moment restriction model is well specified in the sense that there exists a true parameter value θ_0 satisfying the moment condition

$$E\left(\psi_i(\theta_0)\right) = 0, \tag{1}$$

where $\psi_i(\theta) \equiv \psi(x_i, \theta)$.

In such a moment condition model, the most popular estimator is the efficient two-step GMM estimator proposed by Hansen (1982). Let $\bar{\psi}(\theta) = \sum_{i=1}^{n} \psi(x_i, \theta)/n$, $\Omega_n(\theta) = \sum_{i=1}^{n} \psi_i(\theta)\psi_i'(\theta)/n$ and also, let $\tilde{\theta}$ be some first step preliminary (possibly asymptotically inefficient) GMM estimator of θ. The efficient two-step GMM estimator, $\hat{\theta}$ is defined by

$$\hat{\theta} \equiv \arg\min_{\theta \in \Theta} \bar{\psi}'(\theta)\Omega_n^{-1}(\tilde{\theta})\bar{\psi}(\theta).$$

The three-step Euclidean likelihood (3S) estimator as proposed by Antoine, Bonnal and Renault (2007) is considered as computationally less demanding than most of the GMM's alternative estimators including the ETEL estimator. It involves only two quadratic optimization problems which determine the two-step efficient GMM estimator and a GMM first order condition-like solving. To introduce this estimator, let

$$\begin{aligned} J_i(\theta) &= \tfrac{\partial \psi_i(\theta)}{\partial \theta'}, \\ V_n(\theta) &= n^{-1} \textstyle\sum_{i=1}^{n} \psi_i(\theta)(\psi_i(\theta) - \bar{\psi}(\theta))', \\ \pi_i(\theta) &= \tfrac{1}{n} - \tfrac{1}{n}(\psi_i(\theta) - \bar{\psi}(\theta))' V_n^{-1}(\theta)\bar{\psi}(\theta), \\ \bar{G}(\theta) &= \textstyle\sum_{i=1}^{n} \pi_i(\hat{\theta}) J_i'(\theta), \\ \bar{M}(\theta) &= \textstyle\sum_{i=1}^{n} \pi_i(\hat{\theta})\psi_i(\theta)\psi_i'(\theta), \end{aligned} \tag{2}$$

where $\hat{\theta}$ is the efficient two-step GMM.

The 3S estimator is defined as solution of

$$\bar{G}(\hat{\theta}) \left[\bar{M}(\hat{\theta})\right]^{-1} \bar{\psi}(\theta) = 0. \tag{3}$$

$\{\pi_i(\theta) : i = 1, \dots, n\}$ are the implied probabilities yielded by the quadratic discrepancy, also known as the Euclidean empirical likelihood (EEL), function evaluated at θ (see Antoine, Bonnal and Renault (2007)). Equation (3) is similar to the first order condition giving the GMM estimator where the variance and the Jacobian of $\psi_i(\theta)$ at θ_0 are estimated using $\pi_i(\hat{\theta})'$s as weights and are more efficient than sample means which use uniform weights. This efficiency stems from the fact that the Euclidean likelihood implied probabilities provide population expectation estimates using the overidentifying moment conditions as control variables.

The EL estimator also solves a first order condition similar to Equation (3). See Qin and Lawless (1994) and Theorem 2.3 of Newey and Smith (2004). The main difference is that the implied probabilities here have a close form expression and also the Jacobian and the variance are evaluated at a precalculated parameter value. Clearly, this avoids for (3) some numerical issues. Furthermore, the higher order equivalence between the 3S and the EL shows that this approximation does not alter the advantage expected from the resulting estimator.

However, the 3S estimator could suffer of computational inefficiency due to possibly negative Euclidean likelihood implied probabilities. Nonnegative implied probabilities are desirable to allow for probability interpretation in the usual sense. A large implied probability at a sample value is often interpreted as a large concentration of the fundamental probability distribution in the neighborhood of that value. In that respect, implied probabilities are useful in sampling methods that take advantage from the distribution-related information content of the moment conditions (see Brown and Newey (2002)). In the context of the three-step Euclidean likelihood estimator computation, negative implied probabilities can cause the 3S estimator to be unstable and therefore computationally inefficient. In particular, this affects the accuracy of the Jacobian and/or the variance estimators and therefore makes the resulting 3S estimator behave very poorly in finite sample. This harmful effect of negative implied probabilities appears through our Monte Carlo simulation in Section 4.

The use of the shrinkage factor correction proposed by Antoine, Bonnal and Renault (2007) avoids negative implied probabilities. Because both corrected and non corrected implied probabilities are higher order asymptotically equivalent, the resulting estimators from each of them are asymptotically equivalent at least at the first order. The corrected implied probabilities, $\{\pi_i^c(.) : i = 1, \dots, n\}$, that they propose are defined as convex combination of $\pi_i(.)$ and the uniform weight $1/n$ and are nonnegative by construction

$$\pi_i^c(\theta) = \frac{1}{1+\epsilon_n^0(\theta)}\pi_i(\theta) + \frac{\epsilon_n^0(\theta)}{1+\epsilon_n^0(\theta)}\frac{1}{n}$$

where the shrinkage factor $\epsilon_n^0(\theta)$ is given by

$$\epsilon_n^0(\theta) = -n \min\left[\min_{1\le i\le n} \pi_i(\theta), 0\right]. \tag{4}$$

$\epsilon_n^0(\theta)$ converges in probability to 0 while guaranteeing the nonnegativity of $\pi_i^c(\theta)$.

The use of $\pi_i^c(\theta)$ in (3) as proposed by Antoine, Bonnal and Renault (2007) yields the corrected three-step Euclidean likelihood estimator which is also more stable than the 3S estimator. Nevertheless, in globally misspecified models as we discuss in the next section, this shrinkage coefficient will diverge to infinity (as soon as $\psi_i(\theta)$ has an unbounded support) at an unknown rate. This makes the asymptotic behaviour of the corrected 3S estimator hard to characterize in the case of misspecification.

The *modified* three-step Euclidean likelihood (m3S) estimator that we introduce here is a slight modification of this corrected 3S estimator. By construction, it gives more weight to the shrinkage factor such that its rate of divergence could be lower-bounded in the case of misspecification. This shrinkage factor, $\epsilon_n^1(\theta)$ is given by

$$\epsilon_n^1(\theta) = \sqrt{n}\epsilon_n^0(\theta) \tag{5}$$

and the resulting corrected implied probabilities, $\{\tilde{\pi}_i(.) : i = 1, \ldots, n\}$, are given by

$$\tilde{\pi}_i(\theta) = \frac{1}{1+\epsilon_n^1(\theta)}\pi_i(\theta) + \frac{\epsilon_n^1(\theta)}{1+\epsilon_n^1(\theta)}\frac{1}{n}. \tag{6}$$

We label these new corrected implied probabilities as the *modified* implied probabilities to avoid any confusion of the two sets of corrected implied probabilities. Here also and by definition, $\tilde{\pi}_i(\theta) \geq 0$ for all $i = 1, \ldots, n$. Similarly to $\epsilon_n^0(\theta)$ and by Lemma A.2 in Appendix A, $\epsilon_n^1(\theta)$ converges towards 0 in correctly specified models. However, in contrast to $\epsilon_n^0(\theta)$, $\epsilon_n^1(\theta)$ diverges in globally misspecified models at a well defined minimum-bound rate. This difference is crucial as we will see in Section 3. In particular, the $\sqrt{n}$-convergence and the asymptotic Gaussianity that we derive for the *modified* three-step Euclidean likelihood estimator in misspecified models rely on this minimum-bound rate of divergence for the shrinkage factor.

By analogy to the three-step Euclidean likelihood estimator, let

$$\begin{aligned} \tilde{G}(\theta) &= \textstyle\sum_{i=1}^n \tilde{\pi}_i(\hat{\theta})J_i'(\theta), \\ \tilde{M}(\theta) &= \textstyle\sum_{i=1}^n \tilde{\pi}_i(\hat{\theta})\psi_i(\theta)\psi_i'(\theta), \end{aligned} \tag{7}$$

where $\hat{\theta}$ is the efficient two-step GMM estimator.

The *modified* three-step Euclidean likelihood (m3S) estimator is defined as solution of

$$\tilde{G}(\hat{\theta})\left[\tilde{M}(\hat{\theta})\right]^{-1}\bar{\psi}(\theta) = 0. \tag{8}$$

In well specified models, a condition maintained by Antoine, Bonnal and Renault (2007), $\hat{\theta}^{3s} - \hat{\theta}^{el} = O_p(n^{-3/2})$, where $\hat{\theta}^{3s}$ and $\hat{\theta}^{el}$ denote the three-step Euclidean likelihood and the empirical likelihood estimators, respectively. As the ETEL estimator is also proven to be equivalent to EL up to $O_p(n^{-3/2})$, all three share the same $O(n^{-1})$ bias. The following result shows that the *modified* three-step Euclidean likelihood estimator, $\hat{\theta}^{m3s}$, is also higher order equivalent to the empirical likelihood estimator $\hat{\theta}^{el}$. The following assumptions are needed. For brevity, we only highlight in the text those assumptions that are relevant to the exposition and relegate the remainder to the Appendix.

Assumption 2.1 *i)* θ_0 *is an interior point of* Θ*, a compact subset of* $\mathbb{R}^p$*.*
ii) $\psi_i(.)$ *is continuously differentiable in a neighborhood* $\mathcal{N}$ *of* θ_0*.*
iii) $E\left(\psi_i(\theta)\right) = 0 \Leftrightarrow \theta = \theta_0$*.*

iv) $\Omega(\theta_0) = E\left(\psi_i(\theta_0)\psi_i'(\theta_0)\right)$ *is a nonsingular matrix.*

v) $J_0 = E\left(\partial\psi_i(\theta_0)/\partial\theta'\right)$ *is of rank* p*.*

vi) $J_0'\Omega^{-1}(\theta_0)E\left(\psi_i(\theta)\right) = 0 \Leftrightarrow \theta = \theta_0$*.*

vii) The modified three-step Euclidean likelihood estimator is well defined, i.e., there is a sequence $\{\hat{\theta}_{n=1}^{\infty}\}$ *that solves (8) a.s.*

viii) $E\left(\sup_{\theta\in\Theta}\|\psi_i(\theta)\|^{\alpha}\right) < \infty$ *for some* $\alpha > 2$ *and* $E\left(\sup_{\theta\in\mathcal{N}}\|\partial\psi_i(\theta)/\partial\theta'\|\right) < \infty$*.*

Assumption 2.1 provides sufficient conditions for consistency and asymptotic normality of both the efficient two-step GMM estimator $\hat{\theta}$ and the empirical likelihood estimator $\hat{\theta}^{el}$. Assumption 2.1-*(vi)* is an identification condition ensuring the consistency of both $\hat{\theta}^{3s}$ and $\hat{\theta}^{m3s}$.

Theorem 2.1 *If Assumption 2.1, and Assumption A.1 in Appendix A hold, then*

(i) **[Convergence]** $\hat{\theta}^{m3s} \xrightarrow{p} \theta_0$.

(ii) **[Asymptotic normality]** $\sqrt{n}(\hat{\theta}^{m3s} - \theta_0) \xrightarrow{d} \mathcal{N}\left(0, (J_0'[\Omega(\theta_0)]^{-1}J_0)^{-1}\right)$.

(iii) **[Higher order equivalence]** $\hat{\theta}^{m3s} - \hat{\theta}^{el} = O_p(n^{-3/2})$.

Proof: See Appendix A.

The details of the proof of Theorem 2.1 are reported in Appendix A. To establish *(iii)*, we show that $\hat{\theta}^{m3s} - \hat{\theta}^{3s} = O_p(n^{-3/2})$ and deduce the stated order of magnitude by relying on the fact that $\hat{\theta}^{3s} - \hat{\theta}^{el} = O_p(n^{-3/2})$. This result, typically shows that the modified three-step Euclidean likelihood estimator has the same first order asymptotic distribution as the empirical likelihood estimator and both have the same $O(n^{-1})$ bias as well.

The next section studies the 3S and the m3S estimators in the case of model misspecification.

3 The limiting behaviour of the 3S and m3S estimators in misspecified models

In this section, we study the behaviour of the three-step Euclidean likelihood (3S) estimator and the *modified* three-step Euclidean likelihood (m3S) estimator in misspecified models. Following Hall (2000), Hall and Inoue (2003) and Schennach (2007), we consider a moment restriction model like the one given by (1) as misspecified, when there is no value of θ at which the population moment condition is satisfied. In the literature, this case is commonly referred to as non-local or global misspecification. Hall and Inoue (2003) study the two-step GMM estimator under global misspecification. In particular, they establish that the two-step GMM estimator is $\sqrt{n}$-convergent and asymptotically Gaussian in the context of cross sectional data. Since the 3S and the m3S estimators depend on the two-step GMM estimator we partially rely on their results.

3.1 Convergence of the GMM, 3S and m3S estimators

The convergence of the GMM, 3S and m3S estimators require some assumptions. As in the last section and for brevity, we only highlight in the text those assumptions that are relevant to the exposition and the remainder can be found in the Appendix.

Assumption 3.1 *$\{x_i : i = 1, \ldots, n\}$ form an i.i.d. sequence.*

Let $\mu(\theta) = E\left(\psi(x_i, \theta)\right)$ and $\psi_i(\theta) = \psi(x_i, \theta)$.

Assumption 3.2 *i) $\mu : \Theta \to \mathbb{R}^q$ such that $\|\mu(\theta)\| > 0$ for all $\theta \in \Theta$.*
ii) W_n is a positive semidefinite matrix that converges in probability to the positive definite matrix of constants W.
iii) (Identification) There exists $\theta_* \in \Theta$ such that $Q_0(\theta_*) < Q_0(\theta)$ for any $\theta \in \Theta \setminus \{\theta_*\}$ where $Q_0(\theta) = \mu'(\theta)W\mu(\theta)$.

As in Hall (2000) and Hall and Inoue (2003), Assumption 3.2-*(i)* captures the global model misspecification. Assumption 3.2-*(iii)* is the identification condition for a misspecified model. It states that the GMM population objective function given by $Q_0(\theta)$ is minimized at only one point, θ_*, in the parameter set Θ. θ_* is often referred to as the pseudo-true parameter value. This characterization of the pseudo-true value of the GMM estimator is analogue to the characterization of the maximum likelihood estimator's pseudo-true value as formulated by White (1982). One can also refer to Schennach (2007) for the characterization of the empirical likelihood and the exponentially tilted empirical likelihood estimators' pseudo-true values. The existence of pseudo-true value for the estimator of interest is paramount for its convergence. In well specified models, the pseudo-true value corresponds to the true parameter value. In particular, θ_* would correspond to the true parameter value θ_0 and $Q_0(\theta_0) = 0$.

Let $\bar{\theta} = \arg\min_{\theta\in\Theta} \bar{\psi}(\theta)' W_n \bar{\psi}(\theta)$ be the GMM estimator defined by the weighting matrix W_n. Under Assumptions 3.1, 3.2 and Assumption C.1 in the Appendix C, Lemma 1 of Hall (2000) applies and $\bar{\theta}$ converges towards θ_*. This result includes the two-step GMM estimator $\hat{\theta}$ under mild further assumptions. The problem that arises with the two-step GMM estimator is that the weighting matrix it uses depends on a first step GMM estimator $\tilde{\theta}$ which is required to converge. Usually, $\tilde{\theta}$ is obtained by a non random positive definite weighting matrix W^1. We introduce in Appendix B the specific regularity conditions that guarantee the convergence and asymptotic normality of both $\tilde{\theta}$ and $\hat{\theta}$. Let θ_* be the probability limit of $\hat{\theta}$.

Like the two-step GMM estimator, the 3S and the m3S estimators also need pseudo-true values as leading parameter values for their asymptotic behaviour. Recalling that these estimators solve (3) and (8), respectively, their pseudo-true values are determined by the solutions of the population version of these equations.

It is easy to see under some mild conditions that $\bar{G}(\hat{\theta}) \xrightarrow{p} G(\theta_*)$ and $\bar{M}(\hat{\theta}) \xrightarrow{p} M(\theta_*)$ with

$$G(\theta) = E(J_i'(\theta)) - Cov\left(\psi_i'(\theta_*)V^{-1}(\theta_*)\mu(\theta_*),\ J_i'(\theta)\right)$$

and

$$M(\theta) = \Omega(\theta) - Cov\left(\psi_i'(\theta_*)V^{-1}(\theta_*)\mu(\theta_*),\ \psi_i(\theta)\psi_i'(\theta)\right),$$

where $V(\theta) = Var(\psi_i(\theta))$ and $\Omega(\theta) = E(\psi_i(\theta)\psi_i'(\theta))$. The population counterpart of (3) therefore is

$$G(\theta_*)[M(\theta_*)]^{-1}E(\psi_i(\theta)) = 0.$$

If this equation is solved at a single point, $\theta_{**} \in \Theta$, this would be the pseudo-true value of the 3S estimator and one could discuss the asymptotic behaviour of the 3S estimator around this value. We will maintain the existence of such a solution in the next assumption.

As for the m3S estimator, the characterization of the pseudo-true value is made a bit more difficult by the discontinuity of the shrinkage factor $\epsilon_n^1(\hat{\theta})$. In well specified models, the shrinkage factor is meant to vanish asymptotically as confirmed by Lemma A.2. However, in misspecified models and as pointed out by Schennach (2007), it does not vanish. This is the case for $\epsilon_n^0(\hat{\theta})$ and in particular for $\epsilon_n^1(\hat{\theta})$. Actually, if $\psi_i(\theta_*)$ does not have a bounded support, one can establish that these shrinkage factors could diverge to infinity with probability one. In this case, let us consider the following alternative expression of $\tilde{\pi}_i(\theta)$

$$\tilde{\pi}_i(\theta) = \frac{1}{n} - \frac{1}{1+\epsilon_n^1(\theta)}\frac{1}{n}\left(\psi_i(\theta) - \bar{\psi}(\theta)\right)' V_n^{-1}(\theta)\bar{\psi}(\theta). \tag{9}$$

Let $f(x)$ be a real-valued random function. We have

$$\begin{aligned}\sum_{i=1}^n \tilde{\pi}_i(\hat{\theta})f(x_i) \quad &= \quad \tfrac{1}{n}\sum_{i=1}^n f(x_i) - \tfrac{1}{1+\epsilon_n^1(\hat{\theta})}\times \\ &\left(\tfrac{1}{n}\sum_{i=1}^n f(x_i)\psi_i'(\hat{\theta})V_n^{-1}(\hat{\theta})\bar{\psi}(\hat{\theta}) - \bar{\psi}'(\hat{\theta})V_n^{-1}(\hat{\theta})\bar{\psi}(\hat{\theta})\tfrac{1}{n}\sum_{i=1}^n f(x_i)\right).\end{aligned} \tag{10}$$

Under some mild conditions, the terms in the large parenthesis in (10) is asymptotically bounded in probability and since $\epsilon_n^1(\hat{\theta})$ diverges to ∞ with probability one, (10) implies that

$$\sum_{i=1}^n \tilde{\pi}_i(\hat{\theta})f(x_i) = \frac{1}{n}\sum_{i=1}^n f(x_i) + o_p(1).$$

Hence, the population analogue of (8) is

$$E(J_i'(\theta_*))[\Omega(\theta_*)]^{-1}E(\psi_i(\theta)) = 0.$$

If this equation is solved at a unique point of the parameter space Θ, say θ_{**}, this would be the pseudo-true value of the m3S estimator. It is obvious that the m3S and the 3S estimators do not share the same pseudo-true values.

Lemmas C.1 and C.2 in Appendix C discuss more explicitly the conditions that ensure the divergence of the shrinkage factors $\epsilon_n^0(\hat{\theta})$ and $\epsilon_n^1(\hat{\theta})$.

Let $\bar{\mathcal{N}}_*$ be a closed neighbourhood θ_* included in Θ and

$$l_i = \inf_{\theta \in \bar{\mathcal{N}}_*} \psi_i'(\theta) V^{-1}(\theta)\mu(\theta).$$

The absolute continuity of the Lebesgue measure on the real line with respect to the probability distribution of l_i is the most crucial condition making the shrinkage factors diverge to infinity. This condition is rather mild as it includes the Gaussian distribution for instance.

The next results establish the convergence of the 3S and m3S estimators $\hat{\theta}^{3s}$ and $\hat{\theta}^{m3s}$. We make the following useful assumptions.

Assumption 3.3 *i)* $M(\theta_*)$ *is nonsingular and for* $\theta \in \Theta$, $\left(G(\theta_*)[M(\theta_*)]^{-1}\mu(\theta) = 0 \Leftrightarrow \theta = \theta_{**}\right)$.
ii) The three-step Euclidean likelihood estimator is well defined, i.e., there is a sequence $\{\hat{\theta}_n^{3s}\}_{n=1}^{\infty}$ *such that* $\bar{G}(\hat{\theta})[\bar{M}(\hat{\theta})]^{-1}\bar{\psi}(\hat{\theta}^{3s}) = 0$ *a.s.*

Assumption 3.4 *i)* $\forall a, b \in \mathbb{R}$, $a < b$, $\text{Prob}\,(l_i \in (a, b)) \neq 0$.
ii) $\Omega(\theta_*)$ *is nonsingular and, for* $\theta \in \Theta$, $\left(E(J_i'(\theta_*))[\Omega(\theta_*)]^{-1}\mu(\theta) = 0 \Leftrightarrow \theta = \theta_{**}\right)$.
iii) The modified three-step Euclidean likelihood estimator is well defined, i.e., there is a sequence $\{\hat{\theta}_n^{m3s}\}_{n=1}^{\infty}$ *such that* $\tilde{G}(\hat{\theta})[\tilde{M}(\hat{\theta})]^{-1}\bar{\psi}(\hat{\theta}^{m3s}) = 0$ *a.s.*

Assumption 3.3-*(i)* is the identification condition for misspecified model for the 3S estimator problem. Typically, it states that the population version of Equation (3) has a unique solution, θ_{**}, in the parameter set Θ. As previously discussed, θ_{**} is the pseudo-true value for the three-step Euclidean likelihood estimator $\hat{\theta}^{3s}$.

Assumption 3.4-*(i)* means that the Lebesgue measure is absolutely continuous with respect to l_i's probability distribution and implies for l_i to have the whole real line as its distribution's support. Even though this condition could be weakened, it is not too restrictive either as it includes a broad range of probability distributions. As already discussed, this assumption along with some regularity conditions guarantees the shrinkage factor $\epsilon_n^1(\hat{\theta})$ to diverge to infinity. The resulting identification condition is given by Assumption 3.4-*(ii)* with the pseudo-true value of the m3S estimator being θ_{**}.

Obviously, in both cases, θ_{**} depends on both the GMM pseudo-true value θ_* and the asymptotic weighting matrix W. However, we will not explicitly mention this dependence for sake of simplicity. The following remarks present some analogy between the interpretation of the pseudo-true values in the moment condition models that we study here and the fully parametric models.

Remark 1: The norm of the population mean $\mu(\theta)$ evaluates the intensity of model misspecification at the parameter value θ. By definition, the GMM pseudo-true value is

$$\theta_* \equiv \arg\min_{\theta} \|\mu(\theta)\|_V^2,$$

where $\|x\|_V^2 = x'Vx$ and V is a symmetric positive-definite matrix, the so-called GMM norm. Hence θ_* is the parameter value that minimizes the intensity of model misspecification. Of course, a different choice of V points to a different pseudo-true value. There is a parallel between the GMM pseudo-true value in moment condition models and the maximum likelihood (ML) pseudo-true value in fully parametric models. While the GMM pseudo-true value minimizes the intensity of model misspecification, it is well-known that the ML pseudo-true value minimizes the ignorance about the true parametric structure as measured by the Kulback-Leibler divergence (KLIC) of the assumed distribution from the true distribution of the data. What is noticeable in both frameworks is that the GMM pseudo-true value depends on the norm V as the ML pseudo-true value depends on the postulated distribution. For instance, if $\theta = Ex$ is the parameter of interest and one chooses to estimate θ assuming that x is normally distributed, this would lead to the Gaussian pseudo-true value $\theta_*^{(1)}$ minimizing the KLIC between the true distribution of x and the postulated Gaussian distribution. If we rather assume that x follows a Gamma distribution, the corresponding pseudo-true value $\theta_*^{(2)}$ will very likely be different from $\theta_*^{(1)}$.

Remark 2: The fact that the 3S and m3S estimators are defined by the GMM first order local optimality condition makes their pseudo-true values' less obvious to interpret. In general, we can retain that these pseudo-true values are tilted to the GMM pseudo-true value and are also set to minimize the intensity of model misspecification. To see this, let us first observe that $\mu(\theta_{**})$ is defined such that its p components in the space spanned by the columns of $(\Omega(\theta_*))^{-1}J(\theta_*)$ (or $(M(\theta_*))^{-1}G'(\theta_*)$) are all null. Moreover the following expansion holds under mild conditions :

$$J'(\theta_*)\Omega^{-1}(\theta_*)\mu(\theta_{**}) = 0 = J'(\theta_*)\Omega^{-1}(\theta_*)\mu(\theta_*) + J'(\theta_*)\Omega^{-1}(\theta_*)J(\theta_*)(\theta_{**} - \theta_*) + o(\|\theta_{**} - \theta_*\|)$$

hence

$$\theta_{**} - \theta_* = -\left(J'(\theta_*)\Omega^{-1}(\theta_*)J(\theta_*)\right)^{-1} J'(\theta_*)\Omega^{-1}(\theta_*)\mu(\theta_*) + o(\|\theta_{**} - \theta_*\|).$$

Since θ_* makes $\mu(\theta_*)$ small, we also expect $\theta_{**} - \theta_*$ to be small and, by continuity of the function $\mu(\cdot)$, $\mu(\theta_{**})$ is expected to be close to $\mu(\theta_*)$.

Remark 3: As pointed out by one referee, the identification condition introduced by Assumptions 3.3-*(i)* and 3.4-*(i)* may be too restrictive in the sense that, instead of a single value in the parameter space, a family of parameter values, $\{\theta_{**}^{(k)}\}_{k\in I}$ solve the population equation. From the previous remark, the parameter value of interest is the closest to θ_*. The pseudo-true value θ_{**} chosen that way would be estimated by the solution of Equation (3) or (8) closest to the GMM estimator $\hat{\theta}$. As long as the index set I is either finite or discrete, the asymptotic theory that we propose in this paper holds. The case where I is continuous is beyond the scope of this paper.

In the next two results, we assume that Assumption 3.2 holds for the two-step GMM estimator $\hat{\theta}$. The following Theorem establishes the convergence of the 3S estimator, $\hat{\theta}^{3s}$, in globally misspecified models.

Theorem 3.1 *If Assumptions 3.1-3.3, and Assumptions C.1-C.2 in Appendix C hold, then* $\hat{\theta}^{3s} \xrightarrow{p} \theta_{**}$.

Proof: See Appendix C.

The convergence of the m3S estimator, $\hat{\theta}^{m3s}$, is stated by the following result.

Theorem 3.2 *If Assumptions 3.1, 3.2, 3.4, and Assumptions C.1-C.2 in Appendix C hold, and that* $\hat{\theta} - \theta_* = O_p(n^{-1/2})$, *where* $\hat{\theta}$ *is the two-step GMM estimator, then* $\hat{\theta}^{m3s} \xrightarrow{p} \theta_{**}$.

Proof: See Appendix C.

Next, we provide the asymptotic distributions of both the three-step Euclidean likelihood and the *modified* three-step Euclidean likelihood estimators in misspecified models. Since these estimators rely on the two-step GMM estimator, the asymptotic distribution derived by Hall and Inoue (2003) for the two-step GMM in misspecified models is useful for our asymptotic theory. We recall their results here that we also specialize for our use.

3.2 Asymptotic distribution of the two-step GMM estimator in misspecified models

The first step GMM estimator $\tilde{\theta}$ solves

$$\frac{\partial \bar{\psi}'}{\partial \theta}(\tilde{\theta})W^1\bar{\psi}(\tilde{\theta}) = 0, \tag{11}$$

where W^1 is, usually, a non-random weighting matrix. Often, in empirical works, the identity matrix is used as weighting matrix. We treat it here as non-random. Under Assumption 3.1 and Assumptions B.1, C.1 as given in Appendices B and C, the results of Hall and Inoue (2003) apply and

$$\tilde{\theta} - \theta_*^1 = O_p(n^{-1/2}),$$

θ_*^1 being the unique solution of the population analogue of Equation (11).

Actually, a simple Taylor expansion of the first order condition in (11) around θ_*^1 yields

$$0 = \frac{\partial \bar{\psi}'}{\partial \theta}(\theta_*^1) W^1 \bar{\psi}(\theta_*^1) + \left[\frac{\partial \bar{\psi}'}{\partial \theta}(\theta_*^1) W^1 \frac{\partial \bar{\psi}}{\partial \theta'}(\theta_*^1) + (\bar{\psi}'(\theta_*^1) W^1 \otimes I_p) \bar{J}^{(2)}(\theta_*^1) \right] (\tilde{\theta} - \theta_*^1) + O_p(n^{-1}), \quad (12)$$

where I_p is the $p \times p$-identity matrix and

$$\bar{J}^{(2)}(\theta) = \frac{\partial}{\partial \theta'} vec \left(\frac{\partial \bar{\psi}}{\partial \theta'}(\theta) \right).$$

Let

$$\begin{array}{rclrcl} \Omega_n(\theta) & = & \sum_{i=1}^n \psi_i(\theta)\psi_i'(\theta)/n, & J^{(2)}(\theta) & = & E\left((\partial/\partial\theta')vec\left(J_i(\theta)\right)\right), \\ \bar{J}(\theta) & = & \partial\bar{\psi}(\theta)/\partial\theta', & \bar{H}_1(\theta) & = & \bar{J}'(\theta)W^1(\bar{J}(\theta) + (\bar{\psi}'(\theta)W^1 \otimes I_p)\bar{J}^{(2)}(\theta), \\ J(\theta) & = & E\left(J_i(\theta)\right), & H_1(\theta) & = & J'(\theta)W^1 J(\theta) + (\mu'(\theta)W^1 \otimes I_p)J^{(2)}(\theta). \end{array}$$

Since $\bar{H}_1(\theta)$ is a quadratic function of sample means, $\bar{H}_1(\theta)$ is $\sqrt{n}$-convergent for its probability limit $H_1(\theta)$ meaning that $\bar{H}_1(\theta) - H_1(\theta) = O_p(n^{-1/2})$. Therefore,

$$\tilde{\theta} - \theta_*^1 = -H_1^{-1}(\theta_*^1)\bar{J}'(\theta_*^1)W^1\bar{\psi}(\theta_*^1) + O_p(n^{-1}). \quad (13)$$

On the other hand, the two-step GMM estimator solves the first order condition

$$\bar{J}'(\hat{\theta})W_n(\tilde{\theta})\bar{\psi}(\hat{\theta}) = 0, \quad (14)$$

where $W_n(\theta) = [\Omega_n(\theta)]^{-1}$. The stochastic nature of the weighting matrix adds a layer of complexity to the expansion of the two-step GMM estimator equation.

We first expand $\Omega_n(\tilde{\theta})$ around θ_*^1 and then we deduce an expansion of $W_n(\tilde{\theta})$. This latter, ultimately allows to get an expansion for $\hat{\theta}$. We have

$$\Omega_n(\tilde{\theta}) = \Omega_n(\theta_*^1) + R_{q,q}\left(\frac{\partial vec[\Omega]}{\partial \theta'}(\theta_*^1)(\tilde{\theta} - \theta_*^1) \right) + O_p(n^{-1}),$$

where $R_{k,l}(X)$ reshapes the kl-vector X into a $k \times l$-matrix, column-wise.

Let

$$\begin{array}{rcl} \psi_i^* & = & \psi_i(\theta_*^1), \\ J^* & = & J(\theta_*^1), \\ \mu^* & = & \mu(\theta_*^1), \\ W^{-1} & = & \Omega(\theta_*^1), \\ \xi_i(\theta_*^1) & = & \psi_i^*\psi_i^{*'} - \Omega(\theta_1^*) - R_{q,q}\left(\frac{\partial vec[\Omega]}{\partial\theta'}(\theta_*^1)H_1^{-1}(\theta_*^1)\left((\bar{J}'(\theta_*^1) - J^{*'})W^1\mu^* + J^{*'}W^1(\bar{\psi}(\theta_*^1) - \mu^*)\right)\right), \\ \xi_{w,i}(\theta_*^1) & = & -W\xi_i(\theta_*^1)W. \end{array}$$

From the expression of $\tilde{\theta} - \theta_*^1$ given by Equation (13) and up to some arrangements, we have

$$\Omega_n(\tilde{\theta}) = W^{-1} + \frac{1}{n}\sum_{i=1}^{n} \xi_i(\theta_*^1) + O_p(n^{-1}).$$

Clearly, $E\left(\xi_i(\theta_*^1)\right) = 0$ and $\sum_{i=1}^{n} \xi_i(\theta_*^1)/n = O_p(n^{-1/2})$. Furthermore,

$$W_n(\tilde{\theta}) - W = \Omega_n^{-1}(\tilde{\theta}) - W = -\Omega_n^{-1}(\tilde{\theta})(\Omega_n(\tilde{\theta}) - W^{-1})W.$$

Thus

$$W_n(\tilde{\theta}) - W = \frac{1}{n}\sum_{i=1}^{n}\{-W\xi_i(\theta_*^1)W\} + O_p(n^{-1})$$

or equivalently,

$$W_n(\tilde{\theta}) - W = \frac{1}{n}\sum_{i=1}^{n} \xi_{w,i}(\theta_*^1) + O_p(n^{-1}). \tag{15}$$

Thanks to Assumption 3.1 and Assumptions B.2, C.1 in Appendix, we can expand the first order condition for $\hat{\theta}$ in (14) as follows

$$0 = \bar{J}'(\theta_*)W_n(\tilde{\theta})\bar{\psi}(\theta_*) + \left(\bar{J}'(\theta_*)W_n(\tilde{\theta})\bar{J}(\theta_*) + (\bar{\psi}'(\theta_*)W_n(\tilde{\theta}) \otimes I_p)\bar{J}^{(2)}(\theta_*)\right)(\hat{\theta} - \theta_*) + O_p(n^{-1}),$$

Let

$$\begin{aligned}
\psi_{*i} &= \psi_i(\theta_*),\\
\mu_* &= E\psi_{*i},\\
J_* &= J(\theta_*),\\
\bar{H}(\theta) &= \bar{J}'(\theta)W_n(\tilde{\theta})\bar{J}(\theta) + (\bar{\psi}'(\theta)W_n(\tilde{\theta}) \otimes I_p)\bar{J}^{(2)}(\theta),\\
H(\theta) &= J'(\theta)WJ(\theta) + (\mu'(\theta)W \otimes I_p)\, J^{(2)}(\theta).
\end{aligned}$$

Here also, because $\bar{H}(\theta)$ is a polynomial function of sample means, $\bar{H}(\theta_*)$ is $\sqrt{n}$-convergence for its probability limit $H(\theta_*)$ meaning that $\bar{H}(\theta_*) - H(\theta_*) = O_p(n^{-1/2})$. Therefore,

$$\hat{\theta} - \theta_* = -H^{-1}(\theta_*)\bar{J}'(\theta_*)W_n(\tilde{\theta})\bar{\psi}(\theta_*) + O_p(n^{-1}).$$

Thus $\hat{\theta} - \theta_*$ can be written

$$\hat{\theta} - \theta_* = -H^{-1}(\theta_*)\left(\left(\bar{J}'(\theta_*) - J_*'\right)W\mu_* + J_*'\left(W_n(\tilde{\theta}) - W\right)\mu_* + J_*'W\left(\bar{\psi}(\theta_*) - \mu_*\right)\right) + O_p(n^{-1}). \tag{16}$$

From Equations (15) and (16), $\hat{\theta} - \theta_*$ is asymptotically equivalent to a linear function of sample means of centered random vectors which are i.i.d as $x_i : i = 1, \ldots, n$. Since these vectors have finite variance, the central limit theorem applies and $\sqrt{n}(\hat{\theta} - \theta_*) = O_p(1)$ and is asymptotically Gaussian. This is a result of Hall and Inoue (2003).

The main reason of this usual Gaussian asymptotic behaviour of the two-step efficient GMM estimator is the cross sectional nature of the random variables as they are assumed to be i.i.d. This result breaks down in the time series context where the lag dependence is not finite and the moment conditions are globally misspecified. In such a case, as shown by Hall and Inoue (2003) (see also Hall (2000)), the optimal weight for the two-step efficient GMM estimator dictates its rate of convergence to the GMM estimator which therefore may no longer be $\sqrt{n}$-convergence or even asymptotically Gaussian.

3.3 Asymptotic distributions of the three-step Euclidean likelihood estimators

In this section, we derive the asymptotic distribution of both the 3S and the m3S estimators under global misspecification. We find that both are $\sqrt{n}$-convergent and asymptotically characterized by a normal distribution. The asymptotic normality of the 3S estimator is not surprising as its estimating equation sets to zero a smooth function of sample means and the efficient two-step GMM estimator. Since the leading term of the expansion of the GMM estimator is asymptotically Gaussian, the Gaussianity of the 3S estimator in global misspecification becomes quite intuitive.

Besides, the estimating equation of the m3S estimators is not a smooth function of sample means. This makes less apparent the reason of its asymptotically Gaussian behaviour. Let us consider again the optimal sample mean given by (10). As already discussed, we can write

$$\sum_{i=1}^{n} \tilde{\pi}_i(\hat{\theta}) f_i = \frac{1}{n}\sum_{i=1}^{n} f_i - \frac{1}{1+\epsilon_n^1(\hat{\theta})} O_p(1) \tag{17}$$

and $\epsilon_n^1(\hat{\theta})$ diverges to infinity as the sample size grows. This means that the leading term in this expansion is the uniform sample average of the f_is. Considering Equation (8), the leading term of the LHS is therefore a smooth function of sample means. But this pattern of the leading term is not sufficient to guarantee the $\sqrt{n}$-convergence of the m3S estimator, solution of (8). One sufficient, but maybe not necessary, condition for $\hat{\theta}^{m3s}$ to be $\sqrt{n}$-convergent is for the remainder in (17) to vanish faster than $1/\sqrt{n}$, that is

$$\frac{\sqrt{n}}{1+\epsilon_n^1(\hat{\theta})} \xrightarrow{p} 0.$$

Since

$$\frac{\sqrt{n}}{1+\epsilon_n^1(\hat{\theta})} = \frac{1}{1/\sqrt{n}+\epsilon_n^0(\hat{\theta})},$$

this condition is fulfilled. This is the motivation of the choice of $\epsilon_n^1(\hat{\theta})$ as shrinkage factor over $\epsilon_n^0(\hat{\theta})$. Of course, the shrinkage $\epsilon_n^0(\hat{\theta})$ may lead to a $\sqrt{n}$-convergent estimator but, as one could expect, the asymptotic properties of the resulting estimator would be much harder to derive. It is also noteworthy that any shrinkage factor $\epsilon_{\alpha,n}(\theta) = n^{\alpha}\epsilon_n^0(\theta)$ with $\alpha \geq 1/2$ would lead to the same simplifications as $\epsilon_n^1(\theta)$ in globally misspecified models without altering the higher order properties in well specified models. Nevertheless, a large α is only useful in misspecified models. In well specified models, a large α could significantly reduce the effect of correction expected in small sample. For this reason, $\alpha = 1/2$ seems to be an appropriate choice.

The three-step Euclidean likelihood estimator $\hat{\theta}^{3s}$ solves (8) and, by the mean value expansion of (8) around θ_{**}, we have

$$\bar{G}(\hat{\theta})\bar{M}^{-1}(\hat{\theta})\bar{J}(\bar{\theta})(\hat{\theta}^{3s} - \theta_{**}) = -\bar{G}(\hat{\theta})\bar{M}^{-1}(\hat{\theta})\bar{\psi}(\theta_{**}), \tag{18}$$

where $\bar{\theta} \in (\hat{\theta}^{3s}, \theta_{**})$.

To show that $\sqrt{n}(\hat{\theta}^{3s} - \theta_{**})$ is asymptotically normally distributed, we just have to show that the RHS of the last equation scaled by $\sqrt{n}$ is asymptotically Gaussian and the multiplying factor of $\hat{\theta}^{3s} - \theta_{**}$ in the LHS is asymptotically non singular. Let D_* be the probability limit of this multiplicative term. Since $\bar{G}(\hat{\theta})\bar{M}^{-1}(\hat{\theta})\bar{\psi}(\theta_{**})$ is a smooth function of sample mean and converges to 0, it is also $\sqrt{n}$-convergent and (18) implies

$$\hat{\theta}^{3s} - \theta_{**} = -D_*^{-1}\bar{G}(\hat{\theta})\bar{M}^{-1}(\hat{\theta})\bar{\psi}(\theta_{**}) + o_p(n^{-1/2}). \tag{19}$$

Let

$$\begin{array}{rclrcl} \mu_{**} & = & E\left(\psi_i(\theta_{**})\right), & m_* & = & M^{-1}(\theta_*), \\ \omega_* & = & \Omega^{-1}(\theta_*), & v_* & = & V^{-1}(\theta_*), \\ \bar{G}_\pi(\theta) & = & \sum_{i=1}^n \pi_i(\theta)J_i'(\theta), & G_\pi(\theta) & = & \text{plim}\bar{G}_\pi(\theta), \\ \bar{M}_\pi(\theta) & = & \sum_{i=1}^n \pi_i(\theta)\psi_i(\theta)\psi_i'(\theta), & M_\pi(\theta) & = & \text{plim}\bar{M}_\pi(\theta). \end{array}$$

Obviously, $\bar{G}_\pi(\hat{\theta}) = \bar{G}(\hat{\theta})$, $\bar{M}_\pi(\hat{\theta}) = \bar{M}(\hat{\theta})$, $G_\pi(\theta_*) = G(\theta_*)$, and $M_\pi(\theta_*) = M(\theta_*)$.

As suggested by the expansion in Equation (C5) in Appendix C, the leading term in the expansion of the RHS of (19) is a linear function of the vector $\bar{\zeta} - \zeta_0$ obtained by stacking all of the following centered sample means

$$\bar{\psi}(\theta^*) - \mu^*,\ \bar{\psi}(\theta_*) - \mu_*,\ \bar{\psi}(\theta_{**}) - \mu_{**},\ vec\left(\bar{J}(\theta_*^1) - J^*\right),\ vec\left(\bar{J}(\theta_*) - J_*\right),\ vec\left(\Omega_n(\theta_*^1) - \Omega(\theta_*^1)\right),$$

$$vec\left(\Omega_n(\theta_*) - \Omega(\theta_*)\right),\ \frac{1}{n}\sum_{i=1}^n \left(\psi_i(\theta_*) \otimes vec(J_i(\theta_*)) - E(\psi_i(\theta_*) \otimes vec(J_i(\theta_*)))\right),\ \text{and}$$

$$\frac{1}{n}\sum_{i=1}^n \left(\psi_i(\theta_*) \otimes vec(\psi_i(\theta_*)\psi_i'(\theta_*)) - E(\psi_i(\theta_*) \otimes vec(\psi_i(\theta_*)\psi_i'(\theta_*)))\right).$$

Let $\Sigma = Var\left(\sqrt{n}(\bar{\zeta} - \zeta_0)\right)$. If Σ is bounded away from infinity, by the central limit theorem, $\bar{\zeta} - \zeta_0$ is asymptotically Gaussian

$$\sqrt{n}(\bar{\zeta} - \zeta_0) \xrightarrow{d} \mathcal{N}(0, \Sigma)$$

and therefore the 3S estimator is also asymptotically Gaussian. The following assumptions summarize the sufficient conditions for this result.

Assumption 3.5 *i)* $\theta_{**} \in Int(\Theta)$.
ii) $D_* = G(\theta_*)M^{-1}(\theta_*)J(\theta_{**})$ *is nonsingular.*
iii) $\Sigma \equiv Var(\sqrt{n}(\bar{\zeta} - \zeta_0)) < \infty$.
iv) There exists a neighbourhood $\mathcal{N}$ *of* θ_{**} *such that* $E\left(\sup_{\theta\in\mathcal{N}} \|J_i(\theta)\|\right) < \infty$.

Assumption 3.5-*(i)* is a usual one and necessary for the validity of the mean-value expansion. Assumption 3.5-*(ii)* is analogue to the first order identification condition in the GMM theory (see Dovonon and Renault (2008)). This condition is necessary to make the first order expansion of the estimating equation sufficient to characterize $\hat{\theta}^{3s} - \theta_{**}$. Assumption 3.5-*(iii)* ensures the applicability of the central limit theorem while Assumption 3.5-*(iv)* is a dominance condition allowing for a uniform convergence of sample mean of $J_i(\theta)$ in a neighbourhood of θ_{**}.

The following result establishes the asymptotic distribution of the 3S estimator in the case of global misspecification.

Theorem 3.3 *If Assumptions 3.1, 3.3, 3.5 and Assumptions B.2, C.1, and C.2 given in Appendices B and C hold, then there exists a matrix A such that*

$$\sqrt{n}\left(\hat{\theta}^{3s} - \theta_{**}\right) \xrightarrow{d} \mathcal{N}\left(0, D_*^{-1}A\Sigma A' D_*^{-1'}\right).$$

If the moment condition is well specified, $\theta_* = \theta_{**}$ *and*

$$D_*^{-1}A\Sigma A' D_*^{-1'} = \left(J_*'[\Omega(\theta_*)]^{-1}J_*\right)^{-1}$$

which is the usual asymptotic variance.

Proof: See Appendix C.

Theorem 3.3 shows that in global misspecification, the 3S estimator stays $\sqrt{n}$-convergent and asymptotically Gaussian. Its asymptotic variance has the usual 'sandwich' form. On the other hand as one can see by collecting the terms in Equations (C5), (C6), and (C7) in Appendix C, the matrix A is a function of

$$\theta_*^1,\ \mu^*,\ J^*,\ W^1,\ \Omega(\theta_*^1),\ H_1(\theta_*^1),\ \theta_*,\ \mu_*,\ J_*,\ H(\theta_*),\ \partial vec[\Omega](\theta_*^1)/\partial\theta', v_*,\ m_*,\ E\left(\psi_i(\theta_*) \otimes vec(J_i(\theta_*))\right),$$
$$E\left(\psi_i(\theta_*) \otimes vec(\psi_i(\theta_*)\psi_i'(\theta_*))\right),\ \partial vec G_\pi(\theta_*)/\partial\theta',\ \partial vec M_\pi(\theta_*)/\partial\theta'.$$

Clearly, A is straightforward to estimate. Except for the two model parameters θ_*^1 and θ_* which are to be estimated by GMM using respectively the weighting matrices W^1 and $\Omega_n^{-1}(\tilde{\theta})$, all the other quantities are population means and can be estimated by sample averages.

This result also suggests that when the moment condition is well specified, the asymptotic distribution is nothing more than the usual one. For this reason, the asymptotic distribution that we derive can be considered as the model misspecification robust asymptotic distribution of the 3S estimator.

Next, we derive the asymptotic distribution of the *modified* three-step estimator $\hat{\theta}^{m3s}$ in misspecified models. From our previous discussion,

$$\sum_{i=1}^{n} \tilde{\pi}_i(\hat{\theta}) J_i'(\hat{\theta}) = \frac{1}{n} \sum_{i=1}^{n} J_i'(\hat{\theta}) + o_p(n^{-1/2}) \quad \text{and} \quad \sum_{i=1}^{n} \tilde{\pi}_i(\hat{\theta}) \psi_i(\hat{\theta}) \psi_i'(\hat{\theta}) = \frac{1}{n} \sum_{i=1}^{n} \psi_i(\hat{\theta}) \psi_i'(\hat{\theta}) + o_p(n^{-1/2}).$$

Hence,

$$\tilde{G}(\hat{\theta}) \tilde{M}^{-1}(\hat{\theta}) \bar{\psi}(\hat{\theta}^{m3s}) = 0 = \bar{J}'(\hat{\theta}) \Omega_n^{-1}(\hat{\theta}) \bar{\psi}(\hat{\theta}^{m3s}) + o_p(n^{-1/2}).$$

By a mean value expansion of $\bar{\psi}(\theta)$ around θ_{**}, we have

$$\bar{J}'(\hat{\theta}) \Omega_n^{-1}(\hat{\theta}) \bar{J}(\bar{\theta})(\hat{\theta}^{m3s} - \theta_{**}) = -\bar{J}'(\hat{\theta}) \Omega_n^{-1}(\hat{\theta}) \bar{\psi}(\theta_{**}) + o_p(n^{-1/2}),$$

where $\bar{\theta} \in (\hat{\theta}^{m3s}, \theta_{**})$. Since $\bar{J}'(\hat{\theta}) \Omega_n^{-1}(\hat{\theta}) \bar{\psi}(\theta_{**})$ is a smooth function of sample means, it is $\sqrt{n}$-convergent towards its probability limit which, thanks to the identification condition, is 0. Let D_*^m be the probability limit of $\bar{J}'(\hat{\theta}) \Omega_n^{-1}(\hat{\theta}) \bar{J}(\bar{\theta})$. Assuming that D_*^m is nonsingular, we have

$$\hat{\theta}^{m3s} - \theta_{**} = -{D_*^m}^{-1} \bar{J}'(\hat{\theta}) \Omega_n^{-1}(\hat{\theta}) \bar{\psi}(\theta_{**}) + o_p(n^{-1/2}).$$

Thus the asymptotic normality of $(\hat{\theta}^{m3s} - \theta_{**})$ hinges on the asymptotic normality of $\bar{J}'(\hat{\theta}) \Omega_n^{-1}(\hat{\theta}) \bar{\psi}(\theta_{**})$. The expansion given by Equation (C8) in Appendix C shows an asymptotic equivalence between this quantity and a linear combination of the vector $\bar{\zeta}^m - \zeta_0^m$ obtained by stacking all of the following centered sample means

$$\bar{\psi}(\theta^*) - \mu^*,\ \bar{\psi}(\theta_*) - \mu_*,\ \bar{\psi}(\theta_{**}) - \mu_{**},\ vec(\bar{J}(\theta_*) - J_*),\ vec(\bar{J}(\theta_*^1) - J^*),\ vec(\Omega_n(\theta_*^1) - \Omega(\theta_*^1)), \text{ and}$$

$$vec(\Omega_n(\theta_*) - \Omega(\theta_*)).$$

Let $\Sigma^m = Var(\sqrt{n}(\bar{\zeta}^m - \zeta_0^m))$. Clearly, if Σ^m is bounded away from infinity,

$$\sqrt{n}(\bar{\zeta}^m - \zeta_0^m) \xrightarrow{d} \mathcal{N}(0, \Sigma^m)$$

and as a result $\hat{\theta}^{m3s} - \theta_{**}$ also is asymptotically Gaussian. The following assumptions set up the sufficient conditions to reach this result.

Assumption 3.6 *i)* $\theta_{**} \in Int(\Theta)$.
ii) $D_*^m = J'(\theta_*)\Omega^{-1}(\theta_*)J(\theta_{**})$ *is nonsingular.*
iii) $\Sigma^m \equiv Var\left(\sqrt{n}(\bar{\zeta}^m - \zeta_0^m)\right) < \infty$.
iv) There exists a neighbourhood, $\mathcal{N}$*, of* θ_{**} *such that,* $E\left(\sup_{\theta\in\mathcal{N}} \|J_i(\theta)\|\right) < \infty$.

Assumption 3.6 is analogue to Assumption 3.5 and plays the same role for the m3S estimator.

Theorem 3.4 *If Assumptions 3.1, 3.3, 3.6 and Assumptions B.2, C.1, and C.2 in Appendices B and C hold, then there exists a matrix* A^m *such that*

$$\sqrt{n}\left(\hat{\theta}^{m3s} - \theta_{**}\right) \xrightarrow{d} \mathcal{N}\left(0, D_*^{m^{-1}} A^m \Sigma^m A^{m'} D_*^{m^{-1'}}\right).$$

If the moment condition is well specified, $\theta_* = \theta_{**}$ *and*

$$D_*^{m^{-1}} A^m \Sigma^m A^{m'} D_*^{m^{-1'}} = \left(J_*'[\Omega(\theta_*)]^{-1} J_*\right)^{-1}$$

which is the usual asymptotic variance.

Proof: See Appendix C.

Like Theorem 3.3, this result shows that in global misspecification, the m3S estimator also stays $\sqrt{n}$-convergent and asymptotically Gaussian. Its asymptotic variance also has the usual ‘sandwich’ form. From the terms given by the expansions in Equation (C8) in Appendix C, A^m is a function of

$$\theta_*^1,\ \mu^*,\ J^*,\ W^1,\ \Omega(\theta_*^1),\ H_1(\theta_*^1),\ \theta_*,\ \mu_*,\ J_*,\ H(\theta_*),\ \partial vec[\Omega](\theta_*^1)/\partial\theta',\ \omega_* \text{ and } J^{(2)}(\theta_*).$$

Similarly to A, A^m is straightforward to estimate. θ_*^1 and θ_* are to be estimated by GMM using respectively the weighting matrices W^1 and $\Omega_n^{-1}(\tilde{\theta})$ and the other quantities can also be estimated as sample means.

This result also suggests that the asymptotic distribution that we derive is the model misspecification robust asymptotic distribution of the m3S estimator. In comparison with the misspecification robust asymptotic distribution of the 3S estimator, it is worth mentioning that this latter is much harder to implement as the matrix A that it depends on is function of many more parameters. In a statistical point of view, the dependence of A on large number of parameters (though easy to calculate) may lead to a larger estimation error in the natural estimator of A.

These results also show that the two estimators that we consider in this paper have very interesting properties with respect to the alternative most useful moment condition-based estimators. In well specified models, they have the same higher order bias as the EL and ETEL estimators while in misspecified models, they stay $\sqrt{n}$-convergent for they pseudo-true values and asymptotically Gaussian as do the ET and ETEL estimators. Moreover, they are computationally more tractable than all of the estimators in the class of minimum discrepancy estimators and the ETEL estimator as well.

4 Simulations

The Monte Carlo experiments in this section, in addition to the evaluation of the effect of the shrinkage on the 3S estimator, illustrates the two main theoretical results of this paper. Namely, the higher order equivalence of the modified three-step (m3S) Euclidean likelihood estimator and the empirical likelihood (EL) estimator in well specified models and the $\sqrt{n}$-convergence of the 3S and m3S estimators in globally misspecified models. In addition to these three estimators, we consider several alternative estimators including the efficient two-step GMM, the Euclidean empirical likelihood (EEL) estimator, the corrected 3S estimator (m3S0) proposed by Antoine, Bonnal and Renault (2007) which uses $\epsilon_n^0(\theta)$ as shrinkage factor, the exponential tilting (ET) estimator of Kitamura and Stutzer (1997), the exponentially tilted empirical likelihood (ETEL) estimator of Schennach (2007).

The prediction from our theory is that the m3S estimator should have a similar finite sample bias as the EL, the 3S, the m3S0 and the ETEL estimators in well specified models and on the other hand the 3S and the m3S should be $\sqrt{n}$-convergent in globally misspecified models. The GMM estimator is considered as a benchmark, the EEL and the ET estimators are considered because they are partially involved in the derivation of 3S, m3S estimators on one hand and the ETEL estimator on the other hand.

Monte Carlo designs

Our Monte Carlo designs are the same as those used by Schennach (2007). The first one, *Design D*, is also used by Hall and Horowitz (1996), Imbens, Spady and Johnson (1998) and Kitamura (2001). This design generates at each replication a sample of n independent copies of $x_i \equiv (x_{i1}, x_{i2})' \sim \mathcal{N}(0, 0.16 I_2)$, $i = 1, \dots, n$, and $(x_{ik} : k = 3, \dots, K)$, where x_{ik} are independent and identically distributed with $x_{ik} \sim \chi_1^2$. The moment condition model that we consider to fit this data, is

$$E\left(\psi(x_i, \theta)\right) = 0 : \quad \psi(x_i, \theta) = \left(r(x_i, \theta), \quad r(x_i, \theta)x_{i2}, \quad r(x_i, \theta)(x_{i3} - 1), \quad \dots, \quad r(x_i, \theta)(x_{iK} - 1)\right)',$$

$$r(x_i, \theta) = \exp(-0.72 - (x_{i1} + x_{i2})\theta + 3x_{i2}) - 1.$$

The unique parameter value solving these moment conditions is $\theta_0 = 3.0$. This moment condition model has the interest of not being linear and also the third moments of the estimating functions are not trivially null. In these two cases, the estimators that we consider are actually trivially higher order equivalent. This design is used for the first illustration in which we vary both K, the number of moment conditions and n, the sample size. The number of replications we consider throughout in these Monte Carlo experiments is 10,000.

The Monte Carlo designs that we consider to illustrate the $\sqrt{n}$-convergence of the 3S and the m3S estimators under global misspecification are the following Designs C and M. *Design C* generates, for each replication, n independent copies of $x_i \sim \mathcal{N}(0,1)$ fitted by the moment condition model

$$E\left(\psi(x_i,\theta)\right) = 0 : \quad \psi(x_i,\theta) = (x_i - \theta, \quad (x_i-\theta)^2 - 1)'.$$

Design M(s) generates, for each replication, n independent copies of $x_i \sim \mathcal{N}(0, s^2)$; with $s = 0.6, 0.8, 1.2, 1.4$ and fits the same moment condition model as in Design C which is therefore misspecified.

The true parameter value in Design C is $\theta_0 = 0$ and the pseudo-true value of all of the considered estimators for Design M is $\theta_* = 0$. We consider $n = 50$, $n = 200$, $n = 1,000$ and $n = 5,000$.

Estimators

The EEL, EL, ET and ETEL estimators are computed by the inner-outer loops optimization described by Kitamura (2006). It consists first on determining the implied probabilities as a function of θ via a first optimization (the inner loop optimization). Then, the discrepancy function is formed as a function of θ which is optimized over the parameter space. This is the outer loop optimization. It is worth mentioning that the inner loop optimization is unnecessary for the EEL estimator since the implied probabilities of the Euclidean likelihood has a close form formula.

We rely on Design D for the comparison of the bias and standard deviation of the estimators under consideration. The interval $[-19.5; 25.5]$ is used as parameter space. This parameter set is quite large since the estimates should be concentrated around the true parameter value 3.0. We can admit a convergence failure of the computation process in the occurrence of corner solution. Even if any estimator's computation fails to converge, we keep the sample and the estimated value for this estimator. By doing so, the simulated bias and standard deviation for this specific estimator are more likely to be under estimated as the upper bound is often reached in the case of non convergence. The experiment is carried out with $K = 2$, 3 and 10 as number of moment restrictions and $n = 50$, 100, 200 and 500 as sample sizes.

Table 1 displays the simulated median, bias, standard deviations, interquartile range and the number of convergence failure of the estimators in various configurations. First, one can notice the large number of convergence failure for the EEL estimator in particular in small samples and also with increasing number of moment restrictions. This does not come as a surprise since the continuously updated GMM of Hansen, Heaton and Yaron (1996) is also known to potentially display several outliers and this estimator is known to be identical to the EEL estimator. The 3S estimator also displays some outliers but much more rarely than the EEL estimator. Note that the number of outliers here increases with smaller sample size and larger number of moment restrictions. We explain the failure of the EEL and the 3S by the fact that the implied probabilities are not internally guaranteed to be non negative. This leads to poor estimates of the Jacobian and the variance of the estimating function and translates into some instability of both estimators. None of the other estimators shows any systematic case of convergence failure. In that respect and by comparing the 3S estimator to the m3S0 and m3S estimators we can conclude that the shrinkage of the implied probabilities helps to increase the computation efficiency of the 3S estimator.

The outliers displayed by the EEL and the 3S estimators make them less efficient and often much more biased than the other estimators. The m3S0, m3S, EL and ETEL estimators tend to have the same bias in moderate size samples. The similarity of the simulated bias of the m3S and EL estimators is a confirmation of our theory. The m3S0 and the m3S estimators even appear to have a smaller bias for $n = 50$ and 100 in this design with the m3S outperforming, in general, the m3S0. It is also clear that the GMM and the ET estimators do not share the same higher order bias as the other estimators since, even for $n = 500$, their simulated biases are significantly different. This confirms the theory of Newey and Smith (2004) namely that the GMM and the ET estimators have sources of higher order bias different from the EL estimator.

This Monte Carlo experiment does not show a clear evidence of the "no-moment" problem as outlined by Guggenberger (2008). The large standard deviations of the 3S and EEL estimators in small samples are down to outliers and seem to match the other estimators' standard deviations as the sample size grows. These outliers result from computation issues and their inflating effect on the standard deviations is confirmed by the interquartile ranges which are of similar magnitude across all the considered estimators.

Finally, all of these estimators seem sensitive to the number of moment conditions as they display a larger amount of bias with increasing model size.

$\sqrt{n}$-convergence and Gaussianity under misspecification

We illustrate the behaviour of the three-step estimators under global misspecification using Designs C and $M(s) : s = 0.6, 0.8, 1.2, 1.4$. The EEL, EL, ET and ETEL estimators are estimated as previously by inner-outer loops optimizations. The true and pseudo-true parameter values being 0, we consider the interval $[-22.5, 22.5]$ as the parameter set for the estimation purpose. We consider as convergence failure the occurrence of corner solutions or the cases where either the inner or the outer loop optimization routine fails. The simulated statistics in Table 2 are calculated without the failed samples. The first-step GMM estimator is calculated with the weighting matrix $W = (u|v)$ with $u = (1, 0)'$ and $v = (0, 2/3)'$. W is so chosen to reduce the weight on higher moments in the GMM objective function.

As reported by Table 2, the EL and ETEL estimators fail to converge in 2.83% of the simulated samples for the design $M(0.6)$ with $n = 50$ and in 0.01% of the simulated samples for the design $M(0.8)$ also with $n = 50$. The failure of the ETEL is clearly related to the failure of its EL step. This shortcoming highlights some critical issue with the computation of EL in misspecified models.

Table 2 displays the simulated standard errors for all of the estimators. In the correctly specified model, the simulated standard errors are the same for all of the estimators. This, once again, confirms that these estimators have the same asymptotic distribution as predicted by the theory. The cumulative distribution functions plotted by Figure 2 also confirm this theoretical result.

For the misspecified models, from our theory, we expect the simulated standard deviations of the 3S and m3S estimators to shrink by approximately 2 from $n = 50$ to $n = 200$ and by approximately $\sqrt{5}$ from $n = 1,000$ to $n = 5,000$. From the standard deviations displayed by Table 2 for $M(s)$, $s = 0.6, 0.8, 1.2, 1.4$, the 3S and m3S estimators have their standard deviations shrinking by approximately 2 from $n = 50$ to $n = 200$ and $\sqrt{5}$ from $n = 200$ to $n = 1,000$ and $n = 1,000$ to $n = 5,000$ as expected. This is a confirmation of our theory. Even though we do not study the behaviour of the EEL estimator in misspecified models, our simulation results suggest that this estimator may stay $\sqrt{n}$-convergent in misspecified models. The same observation is valid for the m3S0 estimator though no asymptotic theory is available for this estimator in the case of global misspecification and its asymptotic distribution robust to global misspecification is not known. The $\sqrt{n}$-convergence of the GMM estimator in this experiment confirms the results of Hall and Inoue (2003).

The results for ET and ETEL estimators confirm the related literature. Their simulated standard deviations seem to shrink by $\sqrt{n_2/n_1}$ as the sample size grows from n_1 to n_2. However, it appears that the design $M(0.6)$ sees the standard deviation of ETEL decrease by a narrower proportion than expected from $n = 1,000$ to $n = 5,000$. This is likely related to some impact of EL which is the poorest for this design.

The result of Schennach (2007) regarding the EL estimator in globally misspecified models is confirmed by the designs $M(0.6)$ and $M(0.8)$. The simulated standard deviation of this estimator clearly fails to shrink with growing sample sizes. From the cumulative distributions showed by Figure

2, one can also notice some distortion of the EL estimator as the sample size grows in misspecified models. It is however worthwhile to mention that the EL behaves seemingly as a $\sqrt{n}$-convergent estimator in the misspecified designs $M(1.2)$ and $M(1.4)$.

5 Conclusion

The three-step Euclidean likelihood estimator and its corrected version as proposed by Antoine, Bonnal and Renault (2007) are computationally appealing and also higher order equivalent to the empirical likelihood estimator in well specified models as their difference is $O_p(n^{-3/2})$.

This paper studies the 3S and the corrected 3S estimators under global misspecification and shows that the 3S estimator remains $\sqrt{n}$-convergent in misspecified models and its asymptotic distribution robust to global misspecification is derived. As for the corrected 3S estimator, it appears more difficult to analyze in global misspecification context because of the lack of smoothness in the shrinkage factor. We propose a slight modification of the shrinkage factor allowing to control its growing rate as it diverges in the case of global misspecification. We label the resulting estimator the *modified* three-step Euclidean likelihood (m3S) estimator. We show that the m3S estimator is also higher order equivalent to the EL estimator in well specified models while staying $\sqrt{n}$-convergent and asymptotically Gaussian in globally misspecified models. Its asymptotic distribution robust to misspecification is also proposed. These properties make the 3S and the m3S estimators very attractive alternative to the exponentially tilted empirical likelihood (ETEL) estimator proposed by Schennach (2007) since they have the same enjoyable properties of this latter in addition to being computationally more convenient.

There are some lines of extension of this work that we plan for future research. The empirical likelihood ratio parameter and specification tests as proposed by Owen (1990) and Qin and Lawless (1994) are known to outperform their existing alternative such as the Hansen's (1982) GMM overidentification test. However, these tests are computationally demanding as they depend on the full derivation of the EL estimator. It could be of some interest to study the properties of these tests when the three-step and *modified* three-step Euclidean likelihood estimators are used instead of the EL estimator. A higher order equivalence between the new tests and their original versions may suggest some computationally more appealing alternative.

Figure 1: Simulated cumulative distribution function of the 3S, m3S and EL estimators. Design D.

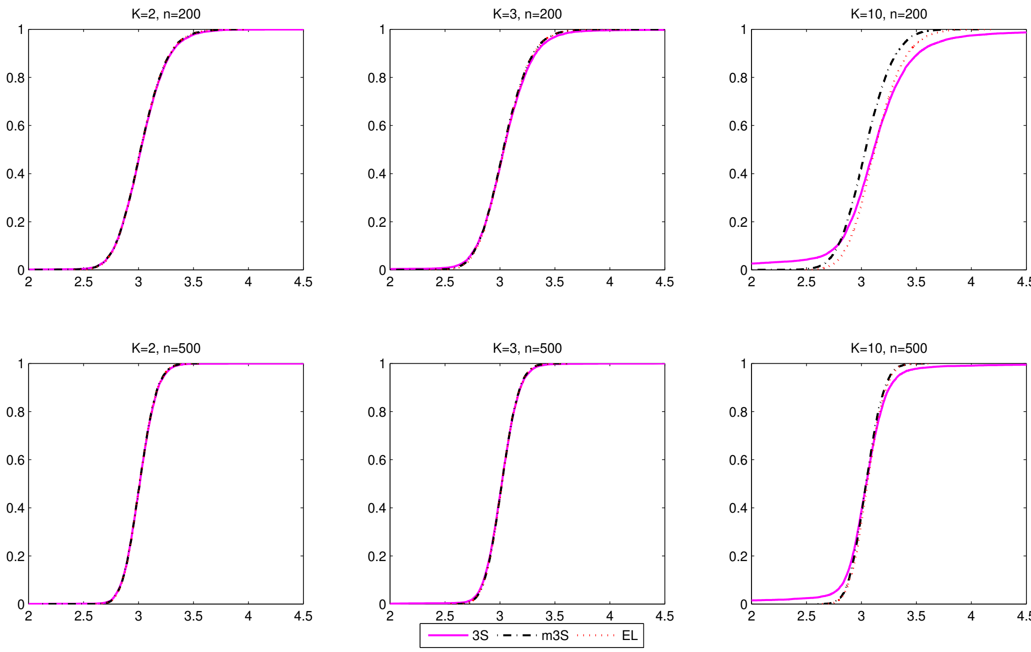

Figure 2: Simulated cumulative distribution function of the 3S, m3S and EL estimators. Well specified vs misspecified models. Designs C and M

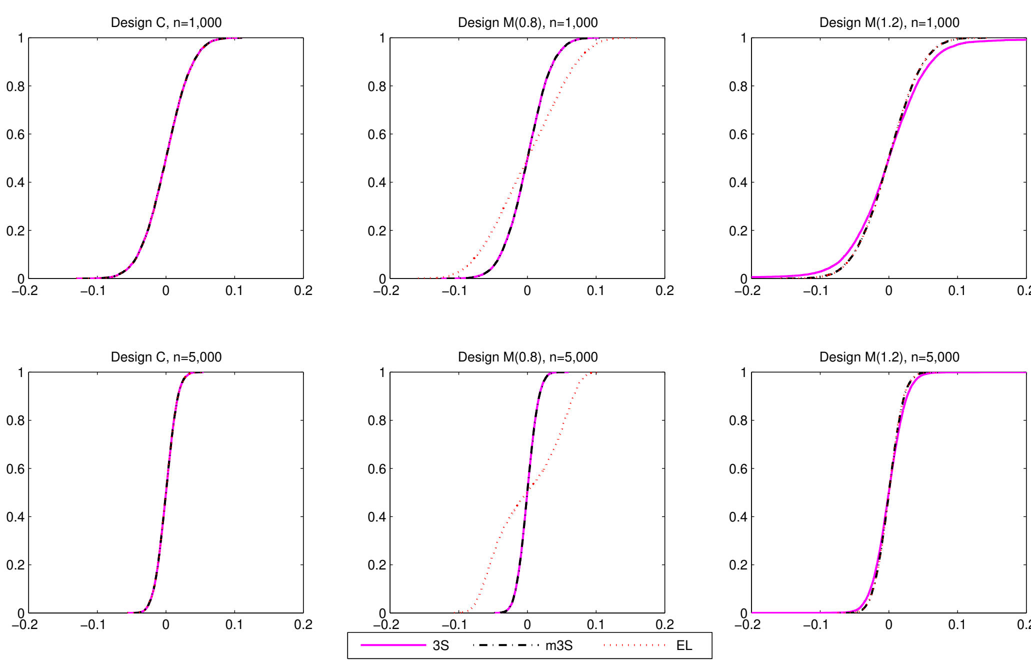

Table 1: The simulated bias, median, standard deviation, and interquartile range of the GMM, EEL, 3S, m3S0, m3S, EL, ETEL and ET estimators from Design D

		GMM	EEL	3S	m3S0	m3S	EL	ETEL	ET
$K = 2, n = 50$	Bias	-0.053	1.340	0.145	0.067	0.034	0.115	0.113	0.163
	Median	2.980	3.133	3.080	3.044	3.031	3.063	3.063	3.089
	Standard deviation	0.600	4.737	0.886	0.517	0.538	0.432	0.430	0.557
	Interquartile range	0.577	0.666	0.582	0.564	0.562	0.545	0.546	0.567
	Convergence failure	0	410	8	0	0	0	0	0
$n = 100$	Bias	-0.001	0.718	0.077	0.058	0.043	0.065	0.065	0.082
	Median	3.014	3.082	3.049	3.042	3.036	3.045	3.044	3.058
	Standard deviation	0.381	3.462	0.698	0.384	0.326	0.284	0.285	0.294
	Interquartile range	0.390	0.419	0.393	0.386	0.382	0.377	0.377	0.383
	Convergence failure	0	207	7	1	0	1	0	0
$n = 200$	Bias	0.013	0.322	0.029	0.032	0.030	0.032	0.032	0.040
	Median	3.008	3.040	3.021	3.021	3.019	3.019	3.020	3.027
	Standard deviation	0.226	2.308	0.457	0.220	0.205	0.199	0.199	0.200
	Interquartile range	0.267	0.276	0.268	0.266	0.265	0.262	0.262	0.263
	Convergence failure	0	92	3	0	0	0	0	0
$n = 500$	Bias	0.009	0.071	0.016	0.014	0.014	0.014	0.014	0.017
	Median	3.007	3.019	3.012	3.012	3.011	3.011	3.011	3.014
	Standard deviation	0.129	0.993	0.316	0.127	0.127	0.126	0.126	0.126
	Interquartile range	0.170	0.171	0.173	0.171	0.170	0.169	0.170	0.170
	Convergence failure	0	18	1	0	0	0	0	0
$K = 3, n = 50$	Bias	-0.098	1.924	0.125	0.081	0.034	0.173	0.166	0.269
	Median	2.935	3.253	3.099	3.061	3.033	3.114	3.112	3.173
	Standard deviation	0.610	5.465	1.296	0.529	0.552	0.464	0.457	0.597
	Interquartile range	0.586	0.887	0.642	0.582	0.577	0.582	0.579	0.634
	Convergence failure	0	492	18	0	0	0	0	0
$n = 100$	Bias	-0.015	1.187	0.067	0.071	0.054	0.090	0.088	0.137
	Median	2.997	3.160	3.065	3.058	3.045	3.066	3.063	3.104
	Standard deviation	0.381	4.316	0.863	0.395	0.328	0.298	0.298	0.327
	Interquartile range	0.390	0.519	0.412	0.393	0.389	0.385	0.385	0.409
	Convergence failure	0	305	9	1	0	0	0	0
$n = 200$	Bias	0.015	0.536	0.052	0.047	0.042	0.048	0.046	0.073
	Median	3.011	3.089	3.035	3.035	3.031	3.037	3.035	3.059
	Standard deviation	0.226	2.864	0.721	0.206	0.207	0.202	0.203	0.211
	Interquartile range	0.270	0.314	0.278	0.270	0.268	0.265	0.266	0.274
	Convergence failure	0	135	8	0	0	0	0	0
$n = 500$	Bias	0.016	0.124	0.012	0.020	0.020	0.021	0.020	0.033
	Median	3.013	3.043	3.017	3.018	3.017	3.018	3.017	3.029
	Standard deviation	0.127	1.220	0.369	0.128	0.128	0.127	0.127	0.129
	Interquartile range	0.170	0.182	0.174	0.171	0.170	0.170	0.170	0.172
	Convergence failure	0	25	2	0	0	0	0	0

Table 1 (Continued): The simulated bias, median, standard deviation, and interquartile range of the GMM, EEL, 3S, m3S0, m3S, EL, ETEL and ET estimators from Design D

		GMM	EEL	3S	m3S0	m3S	EL	ETEL	ET
$K = 10, n = 50$	Bias	-0.627	4.074	-0.449	-0.327	-0.429	0.439	0.319	0.684
	Median	2.466	3.779	2.804	2.745	2.647	3.354	3.248	3.480
	Standard deviation	0.688	7.501	2.445	0.710	0.694	0.630	0.580	1.051
	Interquartile range	0.783	3.273	1.472	0.769	0.731	0.777	0.712	0.977
	Convergence failure	0	856	67	0	0	0	0	0
$n = 100$	Bias	-0.267	4.048	-0.021	0.034	-0.066	0.261	0.201	0.428
	Median	2.766	3.714	3.117	3.038	2.941	3.225	3.172	3.354
	Standard deviation	0.418	7.440	1.768	0.378	0.379	0.362	0.347	0.498
	Interquartile range	0.469	2.007	0.673	0.431	0.427	0.463	0.451	0.557
	Convergence failure	0	890	35	0	0	0	0	0
$n = 200$	Bias	-0.082	2.163	0.117	0.093	0.045	0.137	0.109	0.239
	Median	2.925	3.403	3.113	3.080	3.038	3.121	3.094	3.212
	Standard deviation	0.247	5.597	1.203	0.216	0.217	0.220	0.217	0.260
	Interquartile range	0.295	0.766	0.352	0.282	0.278	0.288	0.283	0.330
	Convergence failure	0	460	17	0	0	0	0	0
$n = 500$	Bias	0.006	0.416	0.029	0.050	0.046	0.058	0.047	0.111
	Median	3.004	3.184	3.045	3.046	3.042	3.054	3.043	3.106
	Standard deviation	0.130	2.036	1.040	0.131	0.129	0.129	0.129	0.139
	Interquartile range	0.174	0.246	0.204	0.178	0.175	0.175	0.175	0.188
	Convergence failure	0	63	16	0	0	0	0	0

Table 2: The simulated bias, median, standard deviation, and interquartile range of the GMM, EEL, 3S, m3S0, m3S, EL, ETEL and ET estimators for Designs C and M

		GMM	EEL	3S	m3S0	m3S	EL	ETEL	ET
Design C	Bias	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$n = 50$	Median	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
	Standard deviation	0.150	0.145	0.147	0.145	0.145	0.146	0.146	0.145
	Interquartile range	0.199	0.194	0.192	0.192	0.192	0.194	0.194	0.195
	Convergence failure	0	0	0	0	0	0	0	0
Design C	Bias	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$n = 200$	Median	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
	Standard deviation	0.072	0.072	0.072	0.072	0.072	0.072	0.072	0.072
	Interquartile range	0.098	0.097	0.097	0.097	0.097	0.097	0.097	0.097
	Convergence failure	0	0	0	0	0	0	0	0
Design C	Bias	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$n = 1000$	Median	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Standard deviation	0.032	0.032	0.032	0.032	0.032	0.032	0.032	0.032
	Interquartile range	0.042	0.042	0.042	0.042	0.042	0.042	0.042	0.042
	Convergence failure	0	0	0	0	0	0	0	0
Design C	Bias	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$n = 5000$	Median	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Standard deviation	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014
	Interquartile range	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019
	Convergence failure	0	0	0	0	0	0	0	0
Design $M(0.6)$	Bias	0.002	0.001	-0.001	0.001	0.002	0.002	0.002	0.001
$n = 50$	Median	0.004	0.000	0.000	0.001	0.002	0.002	0.001	0.001
	Standard deviation	0.167	0.105	0.354	0.210	0.179	0.168	0.168	0.165
	Interquartile range	0.258	0.142	0.442	0.293	0.250	0.236	0.236	0.182
	Convergence failure	0	0	0	0	0	283	282	0
Design $M(0.6)$	Bias	0.000	0.000	0.001	0.000	0.001	0.001	0.001	0.000
$n = 200$	Median	0.001	0.000	0.000	0.000	0.000	0.002	0.002	0.000
	Standard deviation	0.108	0.054	0.141	0.113	0.105	0.135	0.119	0.079
	Interquartile range	0.164	0.073	0.175	0.158	0.145	0.196	0.169	0.108
	Convergence failure	0	0	0	0	0	0	0	0
Design $M(0.6)$	Bias	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000
$n = 1000$	Median	-0.001	0.000	0.000	0.000	0.000	0.001	0.000	0.000
	Standard deviation	0.058	0.024	0.055	0.054	0.052	0.120	0.084	0.044
	Interquartile range	0.082	0.033	0.074	0.073	0.069	0.181	0.120	0.060
	Convergence failure	0	0	0	0	0	0	0	0
Design $M(0.6)$	Bias	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$n = 5000$	Median	0.000	0.000	0.000	0.000	0.000	-0.001	0.000	0.000
	Standard deviation	0.027	0.011	0.025	0.025	0.025	0.113	0.061	0.024
	Interquartile range	0.037	0.014	0.034	0.034	0.034	0.167	0.083	0.033
	Convergence failure	0	0	0	0	0	0	0	0

Table 2 (Continued): The simulated bias, median, standard deviation, and interquartile range of the GMM, EEL, 3S, m3S0, m3S, EL, ETEL and ET estimators for Designs C and M

		GMM	EEL	3S	m3S0	m3S	EL	ETEL	ET
Design $M(0.8)$	Bias	0.002	0.001	-0.001	-0.001	0.000	0.001	0.001	0.001
$n = 50$	Median	0.003	0.001	0.001	0.001	0.001	0.001	0.000	0.000
	Standard deviation	0.173	0.123	0.145	0.137	0.134	0.144	0.141	0.130
	Interquartile range	0.246	0.165	0.173	0.171	0.170	0.193	0.188	0.174
	Convergence failure	0	0	0	0	0	1	1	0
Design $M(0.8)$	Bias	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$n = 200$	Median	0.001	0.000	0.000	0.000	0.000	0.001	0.000	0.000
	Standard deviation	0.098	0.062	0.065	0.065	0.065	0.083	0.076	0.068
	Interquartile range	0.135	0.085	0.085	0.085	0.085	0.115	0.104	0.091
	Convergence failure	0	0	0	0	0	0	0	0
Design $M(0.8)$	Bias	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$n = 1000$	Median	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Standard deviation	0.045	0.028	0.030	0.030	0.030	0.055	0.038	0.031
	Interquartile range	0.060	0.037	0.040	0.040	0.040	0.085	0.052	0.042
	Convergence failure	0	0	0	0	0	0	0	0
Design $M(0.8)$	Bias	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$n = 5000$	Median	0.000	0.000	0.000	0.000	0.000	-0.001	0.000	0.000
	Standard deviation	0.020	0.012	0.013	0.013	0.013	0.052	0.019	0.014
	Interquartile range	0.026	0.016	0.018	0.018	0.018	0.097	0.025	0.019
	Convergence failure	0	0	0	0	0	0	0	0
Design $M(1.2)$	Bias	-0.001	-0.001	0.004	-0.001	-0.001	-0.001	-0.001	-0.001
$n = 50$	Median	-0.001	0.001	0.001	0.001	0.001	0.001	0.000	0.002
	Standard deviation	0.180	0.196	0.299	0.182	0.181	0.182	0.182	0.186
	Interquartile range	0.234	0.256	0.258	0.242	0.239	0.240	0.240	0.244
	Convergence failure	0	0	0	0	0	0	0	0
Design $M(1.2)$	Bias	0.000	0.000	0.003	0.000	0.000	0.000	0.000	0.000
$n = 200$	Median	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
	Standard deviation	0.090	0.097	0.206	0.091	0.090	0.091	0.090	0.092
	Interquartile range	0.119	0.130	0.137	0.122	0.122	0.122	0.122	0.124
	Convergence failure	0	0	0	0	0	0	0	0
Design $M(1.2)$	Bias	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.000
$n = 1000$	Median	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Standard deviation	0.040	0.043	0.093	0.040	0.040	0.040	0.040	0.041
	Interquartile range	0.053	0.058	0.064	0.054	0.054	0.054	0.054	0.055
	Convergence failure	0	0	0	0	0	0	0	0
Design $M(1.2)$	Bias	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$n = 5000$	Median	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Standard deviation	0.018	0.019	0.022	0.018	0.018	0.018	0.017	0.018
	Interquartile range	0.024	0.026	0.028	0.024	0.024	0.024	0.024	0.024
	Convergence failure	0	0	0	0	0	0	0	0

Table 2 (Continued): The simulated bias, median, standard deviation, and interquartile range of the GMM, EEL, 3S, m3S0, m3S, EL, ETEL and ET estimators for Designs C and M

		GMM	EEL	3S	m3S0	m3S	EL	ETEL	ET
Design $M(1.4)$	Bias	-0.001	-0.005	0.019	-0.001	-0.002	-0.002	-0.002	-0.003
$n = 50$	Median	-0.001	-0.004	0.000	-0.001	-0.001	-0.002	-0.001	-0.001
	Standard deviation	0.223	0.284	0.731	0.229	0.225	0.233	0.235	0.249
	Interquartile range	0.282	0.385	0.498	0.303	0.299	0.306	0.307	0.327
	Convergence failure	0	0	0	0	0	0	0	0
Design $M(1.4)$	Bias	0.000	0.000	0.034	0.000	0.000	0.000	0.000	0.000
$n = 200$	Median	0.001	0.001	0.000	0.000	0.001	0.001	0.001	0.001
	Standard deviation	0.113	0.147	0.735	0.113	0.114	0.116	0.116	0.123
	Interquartile range	0.145	0.201	0.535	0.154	0.153	0.157	0.156	0.166
	Convergence failure	0	0	1	0	0	0	0	0
Design $M(1.4)$	Bias	0.000	0.000	0.016	0.000	0.000	0.000	0.000	0.000
$n = 1000$	Median	0.000	0.000	0.000	0.000	0.000	0.000	0.000	-0.001
	Standard deviation	0.050	0.067	0.699	0.050	0.051	0.051	0.051	0.054
	Interquartile range	0.066	0.092	0.490	0.067	0.069	0.070	0.069	0.073
	Convergence failure	0	0	0	0	0	0	0	0
Design $M(1.4)$	Bias	0.000	0.000	0.004	0.000	0.000	0.000	0.000	0.000
$n = 5000$	Median	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Standard deviation	0.022	0.030	0.409	0.022	0.023	0.022	0.022	0.024
	Interquartile range	0.029	0.041	0.250	0.030	0.030	0.030	0.031	0.032
	Convergence failure	0	0	0	0	0	0	0	0

A Proofs of results in Section 2

Assumption A.1 *Let*

$$g_n(\theta) = \bar{G}(\hat{\theta})[\bar{M}(\hat{\theta})]^{-1}\bar{\psi}(\theta)$$

and $\mathcal{N}(\epsilon) = \{\theta : \|\theta - \theta_0\| < \epsilon\}$.

i) For some $\epsilon > 0$, g_n *has partial derivatives* $\mathcal{D}_n(\theta) = \partial g_n(\theta)/\partial\theta'$ *on* $\mathcal{N}(\epsilon)$ *such that, for all* $\delta > 0$,

$$\lim_{\epsilon\to 0}\lim_{n\to\infty} \text{Prob}\left(\sup_{\theta\in\mathcal{N}(\epsilon)} \|\mathcal{D}_n(\theta) - \mathcal{D}_n(\theta_0)\| > \delta\right) = 0.$$

ii) There exists a measurable function $b(x)$ *such that, in a neighbourhood of* θ_0 *and for all* $k, l, r = 1, 2, \ldots, q$, $s = 1, 2, \ldots, p$, $|\psi_k(x,\theta)\psi_l(x,\theta)\psi_r(x,\theta)| < b(x)$, $|\psi_k(x,\theta)(\partial\psi_l(x,\theta)/\partial\theta_s)| < b(x)$ *and* $E(b(x)) < \infty$.

Assumption A.1-*(i)* is an asymptotic continuity condition on the gradient of g_n. This condition is required for the Theorem 1 of Robinson (1988) that we rely on for the proof of Theorem 2.1. The point *(ii)* of the same assumption is the usual dominance conditions for uniform convergence.

Lemma A.1 *Let* h *be a continuous function on a compact set* Θ *such that* $\forall\theta \in \Theta$, $h(\theta) = 0 \Leftrightarrow \theta = \theta_0$. *Let* h_n *be a sequel of functions defined on* Θ *and* $\hat{\theta}_n$ *be a sequel of values in* Θ *such that* $h_n(\hat{\theta}_n) = 0$ *a.s. If* $\sup_{\theta\in\Theta}\|h_n(\theta) - h(\theta)\| \xrightarrow{p} 0$, *then* $\hat{\theta}_n \xrightarrow{p} \theta_0$.

Proof: Let $\mathcal{N}$ be a open neighborhood of θ_0 and $\mathcal{N}^c$ its complement. Since h is continuous on Θ, it is also continuous on $\Theta\cap\mathcal{N}^c$ which is compact. Let $\epsilon = \min_{\theta\in\Theta\cap\mathcal{N}^c}\|h(\theta)\|$. Since $\|h(.)\|$ is continuous on the compact set $\Theta\cap\mathcal{N}^c$, there exists $\theta^* \in \Theta\cap\mathcal{N}^c$ such that $\epsilon = \|h(\theta^*)\|$. Clearly, $\epsilon > 0$ since $\theta^* \neq \theta_0$. On the other hand, for the uniform convergence hypothesis, with probability approaching one, $\|h(\hat{\theta}_n)\| = \|h_n(\hat{\theta}_n) - h(\hat{\theta}_n)\| < \epsilon$. By definition of ϵ, $\hat{\theta}_n \notin \mathcal{N}^c$ and then $\hat{\theta}_n \in \mathcal{N}$ $\square$

Lemma A.2 *If Assumptions 2.1 hold,* $\sqrt{n}\epsilon_n^1(\hat{\theta}) \xrightarrow{p} 0$, *where* $\hat{\theta}$ *is the two-step GMM estimator.*

Proof. We follow similar lines as those of the proof of Theorem 2.2 of Antoine, Bonnal and Renault (2007). Let $Y_i = \sup_{\theta\in\Theta}\|\psi_i(\theta)\|$. Since $E(Y_i^\alpha) < \infty$ for $\alpha > 2$, $Var(Y_i)$ is bounded. Therefore, By Lemma 4 of Owen (1990) and also Lemma D.2. of Kitamura, Tripathi and Ahn (2004),

$$\frac{1}{\sqrt{n}}\max_{1\le i\le n} Y_i = o_p(1).$$

One the other hand, since $\hat{\theta}$ is the two-step GMM estimator, $\hat{\theta}-\theta_0 = O_p(n^{-1/2})$. From our dominance conditions and the central limit theorem, a simple mean value expansion allows to deduce that

$$\bar{\psi}(\hat{\theta}) = O_p(n^{-1/2}). \tag{A1}$$

Now, we show that $\min_{1\leq i\leq n}\pi_i(\hat\theta)\geq 0$ with probability approaching one as n grows. This amounts to showing that $\min_{1\leq i\leq n} n\pi_i(\hat\theta)\geq 0$ with probability approaching one as n grows. let $\delta>0$.

$$\begin{aligned}
\text{Prob}\left(\min_{1\leq i\leq n} n\pi_i(\hat\theta)>1-\delta\right) &= \text{Prob}\left(\min_{1\leq i\leq n}\left\{1-\bar\psi'(\hat\theta)V_n^{-1}(\hat\theta)(\psi_i(\hat\theta)-\bar\psi(\hat\theta))\right\}>1-\delta\right)\\
&= 1-\text{Prob}\left(\exists i=1,\dots,n: 1-\bar\psi'(\hat\theta)V_n^{-1}(\hat\theta)(\psi_i(\hat\theta)-\bar\psi(\hat\theta))<1-\delta\right)\\
&= 1-\text{Prob}\left(\exists i=1,\dots,n: \bar\psi'(\hat\theta)V_n^{-1}(\hat\theta)(\psi_i(\hat\theta)-\bar\psi(\hat\theta))>\delta\right)\\
&\geq 1-\text{Prob}\left(\{|\bar\psi'(\hat\theta)V_n^{-1}(\hat\theta)\bar\psi(\hat\theta)|+\|\sqrt{n}\bar\psi(\hat\theta)\|\|V_n^{-1}(\hat\theta)\|\tfrac{1}{\sqrt{n}}\max_{1\leq i\leq n}Y_i\}>\delta\right),
\end{aligned}$$

where $\|X\|=\sqrt{tr(XX')}$. By the uniform dominance conditions in Assumption 2.1-*(viii)* and Lemma 4.3 of Newey and McFadden (1994), $V_n(\hat\theta)\xrightarrow{p}\Omega(\theta_0)$, nonsingular and therefore $V_n^{-1}(\hat\theta)=O_p(1)$. The last inequality implies that

$$\text{Prob}\left(\min_{1\leq i\leq n} n\pi_i(\hat\theta)>1-\delta\right)\geq 1-\text{Prob}\left(o_p(1)>\delta\right).$$

This shows in particular that, as the sample size grows, $\min_{1\leq i\leq n} n\pi_i(\hat\theta)\geq 0$ with probability approaching one. Thus Prob$(\min_{1\leq i\leq n}\pi_i(\hat\theta)\geq 0)\to 1$ as $n\to\infty$ or equivalently, for any $\delta>0$, there exists $n_0\geq 0$ such that for any $n\geq n_0$, Prob$(\min_{1\leq i\leq n}\pi_i(\hat\theta)\geq 0)\geq 1-\delta$. As a result, $\forall\delta>0$, Prob$(\epsilon_n^0(\hat\theta)=0)\geq 1-\delta$. Thus $\forall\delta>0$, Prob$(n\epsilon_n^0(\hat\theta)=0)\geq 1-\delta$. In other words, Prob$(\sqrt{n}\epsilon_n^1(\hat\theta)=0)\to 1$ as $n\to\infty$. Therefore,

$$\sqrt{n}\epsilon_n^1(\hat\theta)\xrightarrow{p}0\Box \tag{A2}$$

Lemma A.3 *Let $f(x,\theta)$ be a measurable $\mathbb{R}^{L\times M}$-valued function of the random variable x and continuous at θ_0 with probability one and let $\hat\theta$ be the two-step GMM estimator. If Assumption 2.1 holds, and there exists a neighborhood $\mathcal{N}(\theta_0)$ of θ_0 included in Θ such that $E\left(\sup_{\theta\in\mathcal{N}(\theta_0)}\|\psi_i(\theta)\|\|f(x_i,\theta)\|\right)<\infty$ and $E\left(\sup_{\theta\in\mathcal{N}(\theta_0)}\|f(x_i,\theta)\|\right)<\infty$ then $\sum_{i=1}^n\tilde\pi_i(\hat\theta)f(x_i,\hat\theta)\xrightarrow{p}E\left(f(x_i,\theta_0)\right)$.*

Proof. It is straightforward that

$$\begin{aligned}
\sum_{i=1}^n\tilde\pi_i(\hat\theta)f_{lm}(x_i,\hat\theta) &= \tfrac{1}{n}\sum_{i=1}^n f_{lm}(x_i,\hat\theta)-\tfrac{1}{1+\epsilon_n^1(\hat\theta)}\Big(\bar\psi'(\hat\theta)V_n^{-1}(\hat\theta)\tfrac{1}{n}\sum_{i=1}^n\psi_i(\hat\theta)f_{lm}(x_i,\hat\theta)\\
&\qquad -\bar\psi'(\hat\theta)V_n^{-1}(\hat\theta)\bar\psi(\hat\theta)\tfrac{1}{n}\sum_{i=1}^n f_{lm}(x_i,\hat\theta)\Big),
\end{aligned}\tag{A3}$$

where $f_{lm}(.,.)$ if the (l,m)-component of $f(.,.)$. By Lemma A.2, $\epsilon_n^1(\hat\theta)\xrightarrow{p}0$. Moreover, applying Lemma 4.3 of Newey and McFadden (1994), we have $\bar\psi'(\hat\theta)\xrightarrow{p}0$, $V_n(\hat\theta)\xrightarrow{p}\Omega(\theta_0)$, $\frac{1}{n}\sum_{i=1}^n\psi_i(\hat\theta)f_{lm}(x_i,\hat\theta)\xrightarrow{p}E\left(\psi_i(\theta_0)f_{lm}(x_i,\theta_0)\right)<\infty$, and $\frac{1}{n}\sum_{i=1}^n f(x_i,\hat\theta)\xrightarrow{p}E\left(f(x_i,\theta_0)\right)\Box$

Proof of Theorem 2.1. (i)**[Convergence]** Let $Z_n=\tilde G(\hat\theta)\left[\tilde M(\hat\theta)\right]^{-1}$ and $\tilde g_n(\theta)=Z_n\bar\psi(\theta)$. From Lemma A.3, $\tilde G(\hat\theta)\xrightarrow{p}J_0'$ and $\tilde M(\hat\theta)\xrightarrow{p}\Omega(\theta_0)$. Let $Z_0=J_0'[\Omega(\theta_0)]^{-1}$ and $g(\theta)=Z_0E\left(\psi_i(\theta)\right)$. By the identification conditions of Assumption 2.1-*(vi)*, $g(\theta)=0$ only at θ_0. Moreover,

$$\tilde g_n(\theta)-g(\theta)=Z_n(\bar\psi(\theta)-E\psi_i(\theta))+(Z_n-Z_0)E\psi_i(\theta).$$

By the Cauchy-Schwarz inequality,

$$\|\tilde{g}_n(\theta) - g(\theta)\| \leq \|Z_n\| \sup_{\theta\in\Theta} \|\bar{\psi}(\theta) - E\psi_i(\theta)\| + \|Z_n - Z_0\| E \sup_{\theta\in\Theta} \|\psi_i(\theta)\|.$$

From Lemma 2.4 of Newey and McFadden (1994), $\sup_{\theta\in\Theta} \|\bar{\psi}(\theta) - E\psi_i(\theta)\| \xrightarrow{p} 0$. In addition, $Z_n - Z_0 \xrightarrow{p} 0$ and we deduce that $\sup_{\theta\in\Theta} \|g_n(\theta) - g(\theta)\| \xrightarrow{p} 0$. Lemma A.1 therefore applies and $\hat{\theta}^{m3s} \xrightarrow{p} \theta_0$.

(ii) [**Asymptotic normality**] $\hat{\theta}^{m3s}$ solves

$$\tilde{G}(\hat{\theta}) \left[\tilde{M}(\hat{\theta})\right]^{-1} \bar{\psi}(\hat{\theta}^{m3s}) = 0.$$

By a mean-value expansion around θ_0,

$$0 = \tilde{G}(\hat{\theta}) \left[\tilde{M}(\hat{\theta})\right]^{-1} \bar{\psi}(\theta_0) + \tilde{G}(\hat{\theta}) \left[\tilde{M}(\hat{\theta})\right]^{-1} \bar{J}(\bar{\theta})(\hat{\theta}^{m3s} - \theta_0),$$

where $\bar{J}(\theta) = (\partial\bar{\psi}(\theta)/\partial\theta')$ and $\bar{\theta} \in (\hat{\theta}^{m3s}, \theta_0)$ and may differ from row to row. Clearly, $\bar{\theta} \xrightarrow{p} \theta_0$ and $\bar{J}(\bar{\theta}) \xrightarrow{p} J_0$ and $\tilde{G}(\hat{\theta}) \left[\tilde{M}(\hat{\theta})\right]^{-1} \bar{J}(\bar{\theta})$ is nonsingular with probability approaching one. Therefore, for n large enough, we have

$$\hat{\theta}^{m3s} - \theta_0 = -\left(\tilde{G}(\hat{\theta}) \left(\tilde{M}(\hat{\theta})\right)^{-1} \bar{J}(\bar{\theta})\right)^{-1} \tilde{G}(\hat{\theta}) \left(\tilde{M}(\hat{\theta})\right)^{-1} \bar{\psi}(\theta_0)$$

thus

$$\hat{\theta}^{m3s} - \theta_0 = -\left(J_0'[\Omega(\theta_0)]^{-1} J_0\right)^{-1} J_0'[\Omega(\theta_0)]^{-1} \bar{\psi}(\theta_0) + o_p(n^{-1/2}).$$

By the central limit theorem $\sqrt{n}\bar{\psi}(\theta_0) \xrightarrow{d} \mathcal{N}(0, \Omega(\theta_0))$. Therefore,

$$\sqrt{n}(\hat{\theta}^{m3s} - \theta_0) \xrightarrow{d} \mathcal{N}\left(0, \left(J_0'[\Omega(\theta_0)]^{-1} J_0\right)^{-1}\right).$$

(iii) [**Higher order equivalence**] To establish that $\hat{\theta}^{m3s} - \hat{\theta}^{el} = O_p(n^{-3/2})$, we show that $\hat{\theta}^{3s} - \hat{\theta}^{m3s} = O_p(n^{-3/2})$ and we use the result of Antoine, Bonnal and Renault (2007), namely that $\hat{\theta}^{3s} - \hat{\theta}^{el} = O_p(n^{-3/2})$ to deduce that $\hat{\theta}^{m3s} - \hat{\theta}^{el} = O_p(n^{-3/2})$. Our proof for $\hat{\theta}^{3s} - \hat{\theta}^{m3s} = O_p(n^{-3/2})$ applies Theorem 1 of Robinson (1988).

By definition, $g_n(\hat{\theta}^{3s}) = 0$ and $\tilde{g}_n(\hat{\theta}^{m3s}) = 0$. From Theorem 4.1 of Antoine, Bonnal and Renault (2007), $\hat{\theta}^{3s} = \theta_0 + o_p(1)$. By the dominance conditions in Assumptions 2.1-*(viii)* and A.1-*(ii)*, $\partial g_n(\theta_0)/\partial\theta' = \mathcal{D}_0 + o_p(1)$, where $\mathcal{D}_0$ is the nonsingular matrix $J_0'\Omega^{-1}(\theta_0)J_0$. Moreover, from (i), $\hat{\theta}^{m3s} = \theta_0 + o_p(1)$. By Theorem 1 of Robinson (1988),

$$\hat{\theta}^{3s} - \hat{\theta}^{m3s} = O_p(\|g_n(\hat{\theta}^{m3s}) - \tilde{g}_n(\hat{\theta}^{m3s})\|). \tag{A4}$$

Hence

$$\hat{\theta}^{3s} - \hat{\theta}^{m3s} \leq O_p\left\{\left\|\bar{G}(\hat{\theta})[\bar{M}(\hat{\theta})]^{-1} - \tilde{G}(\hat{\theta})[\tilde{M}(\hat{\theta})]^{-1}\right\| \|\bar{\psi}(\hat{\theta}^{m3s})\|\right\}$$

$$\leq O_p\left\{\left\|\left(\bar{G}(\hat{\theta}) - \tilde{G}(\hat{\theta})\right)[\tilde{M}(\hat{\theta})]^{-1} - \bar{G}(\hat{\theta})\left([\tilde{M}(\hat{\theta})]^{-1} - [\bar{M}(\hat{\theta})]^{-1}\right)\right\| \|\bar{\psi}(\hat{\theta}^{m3s})\|\right\}.$$

Under our regularity assumptions, $\sum_{i=1}^{n} \pi_i(\hat{\theta}) J_i(\hat{\theta}) \xrightarrow{p} J_0$. Moreover, for any $k = 1, 2, \ldots, q$ and $s = 1, 2, \ldots, p$,

$$\begin{aligned}
\sum_{i=1}^{n} \tilde{\pi}_i(\hat{\theta}) \frac{\partial \psi_{i,k}}{\partial \theta_s}(\hat{\theta}) - \sum_{i=1}^{n} \pi_i(\hat{\theta}) \frac{\partial \psi_{i,k}}{\partial \theta_s}(\hat{\theta}) &= \frac{\epsilon_n^1(\hat{\theta})}{1 + \epsilon_n^1(\hat{\theta})} \left(\frac{1}{n} \sum_{i=1}^{n} \frac{\partial \psi_{i,k}}{\partial \theta_s}(\hat{\theta}) - \sum_{i=1}^{n} \pi_i(\hat{\theta}) \frac{\partial \psi_{i,k}}{\partial \theta_s}(\hat{\theta}) \right) \\
&= \frac{\epsilon_n^1(\hat{\theta})}{1 + \epsilon_n^1(\hat{\theta})} \bar{\psi}'(\hat{\theta}) \left(-V_n^{-1}(\hat{\theta}) \frac{1}{n} \sum_{i=1}^{n} [\psi_i(\hat{\theta}) - \bar{\psi}(\hat{\theta})] \frac{\partial \psi_{i,k}}{\partial \theta_s}(\hat{\theta}) \right) \\
&= \epsilon_n^1(\hat{\theta}) \frac{1}{1 + \epsilon_n^1(\hat{\theta})} \bar{\psi}'(\hat{\theta}) \left(-V_n^{-1}(\hat{\theta}) \frac{1}{n} \sum_{i=1}^{n} [\psi_i(\hat{\theta}) - \bar{\psi}(\hat{\theta})] \frac{\partial \psi_{i,k}}{\partial \theta_s}(\hat{\theta}) \right) \\
&= o_p(n^{-1/2}) O_p(1) O_p(n^{-1/2}) O_p(1) = o_p(n^{-1}).
\end{aligned}$$

Thus $\tilde{G}(\hat{\theta}) - \bar{G}(\hat{\theta}) = o_p(n^{-1})$. Similarly, $\tilde{M}(\hat{\theta}) - \bar{M}(\hat{\theta}) = o_p(n^{-1})$.

On the other hand, since $[\tilde{M}(\hat{\theta})]^{-1} - [\bar{M}(\hat{\theta})]^{-1} = -[\bar{M}(\hat{\theta})]^{-1} \left(\tilde{M}(\hat{\theta}) - \bar{M}(\hat{\theta}) \right) [\tilde{M}(\hat{\theta})]^{-1}$, we deduce that $[\tilde{M}(\hat{\theta})]^{-1} - [\bar{M}(\hat{\theta})]^{-1} = o_p(n^{-1})$.

Furthermore, from (ii), $\hat{\theta}^{m3s} - \theta_0 = O_p(n^{-1/2})$ and the usual mean-value expansion ensures that $\bar{\psi}(\hat{\theta}^{m3s}) = O_p(n^{-1/2})$.

Therefore, $\hat{\theta}^{3s} - \hat{\theta}^{m3s} = O_p(n^{-3/2})$. Since, from Antoine, Bonnal and Renault (2007), $\hat{\theta}^{3s} - \hat{\theta}^{el} = O_p(n^{-3/2})$, we also have $\hat{\theta}^{m3s} - \hat{\theta}^{el} = O_p(n^{-3/2})$ $\square$

B Regularity conditions for the first and two-step GMM estimators $\tilde{\theta}$ and $\hat{\theta}$

The first step GMM estimator $\tilde{\theta}$ is defined as

$$\arg\min_{\theta \in \Theta} \bar{\psi}'(\theta) W^1 \bar{\psi}(\theta).$$

The following assumption ensures the convergence and asymptotic normality of $\tilde{\theta}$ in the case of global misspecification as formalized by Assumption 3.2-*(i)*.

Assumption B.1 *i)* $\|\mu(\theta)\| > 0$ *for all* $\theta \in \Theta$.

ii) W^1 *is a symmetric positive definite matrix.*

iii) There exists $\theta_*^1 \in \Theta$ *such that* $Q_0^1(\theta_*^1) < Q_0^1(\theta)$ *for all* $\theta \in \Theta \setminus \{\theta_*^1\}$, *where* $Q_0^1(\theta) = \mu'(\theta) W^1 \mu(\theta)$.

iv) $\theta_*^1 \in Int(\Theta)$.

v) $\psi(x, .)$ *is twice continuously differentiable on* $Int(\Theta)$ *and* $\partial\psi(., \theta)/\partial\theta'$ *and* $(\partial/\partial\theta') vec[\partial\psi(., \theta)/\partial\theta']$ *are measurable for each* $\theta \in Int(\Theta)$.

vi) There exists a measurable function $b_1(x)$ *such that* $|\psi_k(x, \theta)| < b_1(x)$, $|\partial\psi_k(x, \theta)/\partial\theta_s| < b_1(x)$, $|\partial^2\psi_k(x, \theta)/\partial\theta_s\partial\theta_u| < b_1(x)$ *in a neighbourhood of* θ_*^1, *for all* $k = 1, 2, \ldots, q$ *and* $s, u = 1, 2, \ldots, p$ *and* $E\left(b_1(x)^2\right) < \infty$.

vii) $H_1(\theta_*^1) = J'(\theta_*^1) W^1 J(\theta_*^1) - (E\psi_i'(\theta_*^1) W^1 \otimes I_p) J^{(2)}(\theta_*^1)$ *is nonsingular.*

viii) $Var z_{1,i} < \infty$, *where* $z_{1,i} = (\psi_i'(\theta_*^1), vec'\left(\partial\psi_i(\theta_*^1)/\partial\theta'\right))'$.

Assumptions B.1-*(i)*-*(iii)* are stronger than Assumption 3.2 as the weighting matrix here is not random. Under Assumptions 3.1, B.1-*(i)*-*(iii)* and C.1 in Appendix C, we can apply the Lemma 1 of Hall (2000) and deduce that $\tilde{\theta} \xrightarrow{p} \theta_*^1$. Moreover, thanks to Theorem 2 of Hall and Inoue (2003), if Assumptions 3.1, B.1 and C.1 hold, $\sqrt{n}(\tilde{\theta} - \theta_*^1) \xrightarrow{d} \mathcal{N}(0, \omega_1)$. One can refer to Hall and Inoue (2003) for an explicit expression for ω_1. These conditions also imply that $\Omega_n(\tilde{\theta}) = \sum_{i=1}^n \psi_i(\tilde{\theta})\psi_i'(\tilde{\theta})/n$ is convergent for $E\left(\psi_i(\theta_*^1)\psi_i'(\theta_*^1)\right)$. We will explicitly assume next that this probability limit is nonsingular. This additional assumption guarantees the two-step GMM estimator computation in large sample. The regularity conditions for the two-step GMM estimator are presented next.

Assumption B.2 *i) Assumption B.1 holds.*
ii) $\Omega(\theta_*^1)$ *is nonsingular and let* $W = \left(\Omega(\theta_*^1)\right)^{-1}$.
iii) There exists $\theta_* \in \Theta$ *such that* $Q_0(\theta_*) < Q_0(\theta)$ *for all* $\theta \in \Theta \setminus \{\theta_*\}$, *where* $Q_0(\theta) = \mu'(\theta)W\mu(\theta)$.
iv) $\theta_* \in Int(\Theta)$.
v) There exists a measurable function $b_2(x)$ *such that* $|\psi_k(x,\theta)| < b_2(x)$, $|\partial\psi_k(x,\theta)/\partial\theta_s| < b_2(x)$, $|\partial^2\psi_k(x,\theta)/\partial\theta_s\partial\theta_u| < b_2(x)$ *in a neighbourhood of* θ_*, *for all* $k = 1, 2, \dots, q$ *and* $s, u = 1, 2, \dots, p$ *and* $E\left(b_2(x)^2\right) < \infty$.
vi) $H(\theta_*) = J'(\theta_*)WJ(\theta_*) - (E\psi_i'(\theta_*)W^1 \otimes I_p)J^{(2)}(\theta_*)$ *is nonsingular.*
viii) $Varz_{2,i} < \infty$, *where* $z_{2,i} = (\psi_i'(\theta_*), vec'\left(\psi_i(\theta_*^1)\psi_i'(\theta_*^1)\right), vec'\left(\partial\psi_i(\theta_*)/\partial\theta'\right))'$.

Assumptions B.2-*(i)*-*(iii)* are also a particular case of Assumption 3.2. Under Assumptions 3.1, B.2-*(i)*-*(iii)* and C.1, Lemma 1 of Hall (2000) holds and $\hat{\theta} \xrightarrow{p} \theta_*$. Furthermore, from the asymptotic result in Theorem 3 of Hall and Inoue (2003), if Assumptions 3.1, B.2, and C.1 hold, $\sqrt{n}(\hat{\theta} - \theta_*) \xrightarrow{d} \mathcal{N}(0, \omega_2)$. One can refer to Hall and Inoue (2003) for an explicit expression for ω_2.

C Proofs of results in Section 3

Assumption C.1 *i)* Θ *is compact.*
ii) $\psi(., \theta)$ *is measurable for each* $\theta \in \Theta$ *and* $\psi_i(.)$ *is continuous with probability one on* Θ.
iii) $E\left(\sup_{\theta\in\Theta} \|\psi_i(\theta)\|\right) < \infty$.

Assumption C.2 *i)* $\psi(x, .)$ *is differentiable with probability one on* Θ.
ii) There exists a measurable function $b(x)$ *such that, in a neighbourhood of* θ_* *and for all* $k, l, r = 1, 2, \dots, q$, $s = 1, 2, \dots, p$, $|\psi_k(x,\theta)\psi_l(x,\theta)\psi_r(x,\theta)| < b(x)$, $|\psi_l(x,\theta)(\partial\psi_k(x,\theta)/\partial\theta_s)| < b(x)$, $|\partial\psi_k(x,\theta)/\partial\theta_s| < b(x)$ *and* $E(b(x)) < \infty$.

Proof of Theorem 3.1. Under Assumption 3.2, the two-step GMM estimator $\hat{\theta}$ is convergent for θ_* and Assumptions C.1 and C.2 allow Lemma 4.3 of Newey and McFadden (1994) to apply and $\bar{G}(\hat{\theta}) \xrightarrow{p} G(\theta_*)$ and $\bar{M}(\hat{\theta}) \xrightarrow{p} M(\theta_*)$ so that $\bar{G}(\hat{\theta})\bar{M}^{-1}(\hat{\theta}) \xrightarrow{p} G(\theta_*)M(\theta_*)^{-1}$.
Let $h_n(\theta) = \bar{G}(\hat{\theta})\bar{M}(\hat{\theta})^{-1}\bar{\psi}(\theta)$ and $h(\theta) = G(\theta_*)M^{-1}(\theta_*)E(\psi_i(\theta))$, for $\theta \in \Theta$.

By definition, $h_n(\hat{\theta}^{3s}) = 0$ and, by Assumption 3.3 , for $\theta \in \Theta$, $h(\theta) = 0 \Leftrightarrow \theta = \theta_{**}$. To apply the convergence result of Lemma A.1, we need to establish the additional uniform convergence condition $\sup_{\theta\in\Theta} \|h_n(\theta) - h(\theta)\| \xrightarrow{p} 0$.

$$\begin{aligned}
\|h_n(\theta) - h(\theta)\| &= \left\| \bar{G}(\hat{\theta})\bar{M}^{-1}(\hat{\theta})\bar{\psi}(\theta) - G(\theta_*)M^{-1}(\theta_*)E(\psi_i(\theta)) \right\| \\
&= \left\| \left(\bar{G}(\hat{\theta})\bar{M}^{-1}(\hat{\theta}) - G(\theta_*)M^{-1}(\theta_*) \right) (\overline{\psi}(\theta) - E(\psi_i(\theta))) \right. \\
&\quad \left. + \left(\bar{G}(\hat{\theta})\bar{M}^{-1}(\hat{\theta}) - G(\theta_*)M^{-1}(\theta_*) \right) E(\psi_i(\theta)) + G(\theta_*)M^{-1}(\theta_*)(\overline{\psi}(\theta) - E(\psi_i(\theta))) \right\| \\
&\le \left\| \bar{G}(\hat{\theta})\bar{M}^{-1}(\hat{\theta}) - G(\theta_*)M^{-1}(\theta_*) \right\| \left[\|\overline{\psi}(\theta) - E(\psi_i(\theta))\| + \|E(\psi_i(\theta))\| \right] \\
&\quad + \|G(\theta_*)\| \|M(\theta_*)\|^{-1} \|\overline{\psi}(\theta) - E(\psi_i(\theta))\|.
\end{aligned}$$

Clearly and from Assumption C.1, $\sup_{\theta\in\Theta} \|E(\psi_i(\theta))\| \le E(\sup_{\theta\in\Theta} \|\psi_i(\theta)\|) < \infty$. Thanks to the same assumption, we can also apply Lemma 4.2 of Newey and McFadden (1994) and $\sup_{\theta\in\Theta} \|\overline{\psi}(\theta) - E(\psi_i(\theta))\| \xrightarrow{p} 0$. As a result, $\sup_{\theta\in\Theta} \|h_n(\theta) - h(\theta)\| \xrightarrow{p} 0$. Therefore, from Lemma A.1, we can deduce that $\hat{\theta}^{3s} \xrightarrow{p} \theta_{**}$ $\square$

Lemma C.1 *Let $x_i, i = 1, 2, \ldots, n$ be an i.i.d random sample and let $y(x_i, \theta)$ be a measurable real valued function of x_i and θ, continuous with probability one at each $\theta \in \bar{\mathcal{N}}$, where $\bar{\mathcal{N}}$ is a compact subset of Θ. Let $\bar{\theta}$ be a random vector that lies in $\bar{\mathcal{N}}$ with probability approaching one as n grows to infinity.*
If $\text{Prob}[\inf_{\theta\in\bar{\mathcal{N}}} y(x_i, \theta) \in (a, b)] \neq 0$ *for any a and b on the real line such that $a < b$, then for any $M > 0$,* $\text{Prob}\{\max_{1\le i\le n} y(x_i, \bar{\theta}) > M\} \to 1$ *as $n \to \infty$.*

Proof: Because $\bar{\theta} \in \bar{\mathcal{N}}$ with probability approaching one as n grows to infinity, for large n and for any $i = 1, \ldots, n$, $\inf_{\theta\in\bar{\mathcal{N}}} y(x_i, \theta) \le y(x_i, \bar{\theta})$ with probability one. Therefore, with probability approaching one as n grows,

$$\max_{1\le i\le n} \left(\inf_{\theta\in\Theta} y(x_i, \theta) \right) \le \max_{1\le i\le n} y(x_i, \bar{\theta}).$$

Then, for $M > 0$,

$$\text{Prob}\left(\max_{1\le i\le n} \left(\inf_{\theta\in\bar{\mathcal{N}}} y(x_i, \theta) \right) > M \right) \le \text{Prob}\left(\max_{1\le i\le n} y(x_i, \bar{\theta}) > M \right).$$

As $x_i, i = 1, \ldots, n$ are $i.i.d$, so are $\inf_{\theta\in\bar{\mathcal{N}}} y(x_i, \theta), i = 1, \ldots, n$ and hence,

$$\begin{aligned}
\text{Prob}\left(\max_{1\le i\le n} \left(\inf_{\theta\in\bar{\mathcal{N}}} y(x_i, \theta) \right) > M \right) &= 1 - \text{Prob}\left\{ \max_{1\le i\le n} \left(\inf_{\theta\in\bar{\mathcal{N}}} y(x_i, \theta) \right) \le M \right) \\
&= 1 - \text{Prob}\left(\inf_{\theta\in\bar{\mathcal{N}}} y(x_i, \theta) \le M; \forall i = 1, \ldots, n \right) \\
&= 1 - \left(\text{Prob}\left(\inf_{\theta\in\bar{\mathcal{N}}} y(x_1, \theta) \le M \right) \right)^n.
\end{aligned}$$

Since $\text{Prob}\left(\inf_{\theta\in\bar{\mathcal{N}}} y(x_1, \theta) \in (a, b)\right) \neq 0$ for any $a < b$, $0 < \text{Prob}\left(\inf_{\theta\in\bar{\mathcal{N}}} y(x_1, \theta) \le M\right) < 1$.
Thus, $\lim_{n\to\infty} \left(\text{Prob}\left(\inf_{\theta\in\bar{\mathcal{N}}} y(x_1, \theta) \le M\right)\right)^n = 0$. As a result, $\text{Prob}\left(\max_{1\le i\le n} y(x_i, \bar{\theta}) > M\right) \to 1$ as $n \to \infty$
$\square$

Lemma C.2 *Under Assumptions 3.1, 3.4 and C.2, if the GMM estimator $\hat{\theta}$ is such that $\hat{\theta}-\theta_* = O_p(n^{-1/2})$, then* Prob $\left(\epsilon_n^0(\hat{\theta}) > M\right) \to 1$, *for all $M > 0$, where $\epsilon_n^0(\theta) = -n \min[\min_{1\leq i\leq n} \pi_i(\theta), 0]$.*

Proof: By definition, $\epsilon_n^0(\hat{\theta}) = \max\{\max_{1\leq i\leq n} -n\pi_i(\hat{\theta}); 0\}$. As a result, the sets of events $\{\epsilon_n^0(\hat{\theta}) > M \geq 0\}$ and $\{\max_{1\leq i\leq n}\left(-n\pi_i(\hat{\theta})\right) > M\}$ are equal. By the definition of $\pi_i(\theta)$ in Equation (2), this latter can be written

$$\left\{\max_{1\leq i\leq n}\left(\left((\psi_i(\hat{\theta})-\bar{\psi}(\hat{\theta}))'V_n^{-1}(\hat{\theta})\bar{\psi}(\hat{\theta})\right)-1\right) > M\right\}.$$

On the other hand, since $\hat{\theta}$ is convergent for θ_*, by the dominance conditions in Assumption C.2, $\bar{\psi}'(\hat{\theta})V_n^{-1}(\hat{\theta})\bar{\psi}(\hat{\theta})$ converges in probability to a fixed scalar c. Then, to complete the proof, it suffices to show that

$$\text{Prob}\left(\max_{1\leq i\leq n}\psi_i'(\hat{\theta})V_n^{-1}(\hat{\theta})\bar{\psi}(\hat{\theta}) > M\right) \to 1, \quad \forall M.$$

Moreover,

$$\psi_i'(\hat{\theta})V_n^{-1}(\hat{\theta})\bar{\psi}(\hat{\theta}) = \psi_i'(\hat{\theta})\left[V_n^{-1}(\hat{\theta})\bar{\psi}(\hat{\theta}) - V^{-1}(\theta_*)E(\psi_i(\theta_*))\right] + \psi_i'(\hat{\theta})V^{-1}(\theta_*)E(\psi_i(\theta_*)).$$

Because $\hat{\theta}$ is $\sqrt{n}$-convergent, thanks to Assumption C.2, $V_n^{-1}(\hat{\theta})\bar{\psi}(\hat{\theta}) - V^{-1}(\theta_*)E(\psi_i(\theta_*)) = O_p(n^{-1/2})$. Still by Assumption C.2, $E(\sup_{\theta\in\bar{\mathcal{N}}_*}\|\psi_i(\theta)\|^2) < \infty$, where $\bar{\mathcal{N}}_*$ is a closed neighbourhood of θ_* included in Θ. By Lemma 4 in Owen (1990) and Lemma D.2. in Kitamura, Tripathi and Ahn (2004), $\max_{1\leq i\leq n}\sup_{\theta\in\bar{\mathcal{N}}_*}\|\psi_i(\theta)\| = o_p(n^{1/2})$. Hence, for n large enough and by the Cauchy-Schwartz inequality,

$$\begin{aligned}\left|\psi_i'(\hat{\theta})\left(V_n^{-1}(\hat{\theta})\bar{\psi}(\hat{\theta}) - V^{-1}(\theta_*)E(\psi_i(\theta_*))\right)\right| \leq & \frac{1}{\sqrt{n}}\max_{1\leq i\leq n}\sup_{\theta\in\bar{\mathcal{N}}_*}\|\psi_i(\theta)\| \\ & \times\sqrt{n}\|V_n^{-1}(\hat{\theta})\bar{\psi}(\hat{\theta}) - V^{-1}(\theta_*)E(\psi_i(\theta_*))\| = o_p(1).\end{aligned}$$

Then, $\psi_i'(\hat{\theta})\left[V_n^{-1}(\hat{\theta})\bar{\psi}(\hat{\theta}) - V^{-1}(\theta_*)E(\psi_i(\theta_*))\right] = o_p(1)$ uniformly over $i = 1,\ldots,n$. Hence, it suffices to show that Prob $\left(\max_{1\leq i\leq n}\psi_i'(\hat{\theta})V^{-1}(\theta_*)E(\psi_i(\theta_*)) > M\right) \to 1$ as $n\to\infty$, for all M. Thanks to Assumption 3.4, we can apply Lemma C.1 with $y(x_i,\theta) = \psi_i'(\theta)V^{-1}(\theta_*)E(\psi_i(\theta_*))$ and the result follows □

Proof of Theorem 3.2. Let $y(x_i,\theta) = J_i(\theta)$ or $\psi_i(\theta)\psi_i'(\theta)$. By definition,

$$\sum_{i=1}^n\tilde{\pi}_i(\hat{\theta})y(x_i,\hat{\theta}) = \frac{1}{n}\sum_{i=1}^n y(x_i,\hat{\theta}) - \frac{1}{1+\epsilon_n^1(\hat{\theta})}\frac{1}{n}\sum_{i=1}^n\left((\psi_i(\hat{\theta})-\bar{\psi}(\hat{\theta}))'V_n^{-1}(\hat{\theta})\bar{\psi}(\hat{\theta})\right)y(x_i,\hat{\theta}).$$

By Lemma 4.3 of Newey and McFadden (1994), Under some regularity conditions,

$$\frac{1}{n}\sum_{i=1}^n\left((\psi_i(\hat{\theta})-\bar{\psi}(\hat{\theta}))'V_n^{-1}(\hat{\theta})\bar{\psi}(\hat{\theta})\right)y(x_i,\hat{\theta}) = O_p(1).$$

Besides, Lemma C.2 ensures that

$$\frac{1}{1+\epsilon_n^1(\hat{\theta})} = \frac{1}{1+\sqrt{n}\epsilon_n^0(\hat{\theta})} = o_p(1)$$

since $\epsilon_n^0(\hat{\theta})$ diverges to infinity. Therefore, $\sum_{i=1}^n\tilde{\pi}_i(\hat{\theta})y(x_i,\hat{\theta}) = \sum_{i=1}^n y(x_i,\hat{\theta})/n + o_p(n^{-1/2})$.

Specifically,

$$\sum_{i=1}^n\tilde{\pi}_i(\hat{\theta})J_i(\hat{\theta}) = \frac{1}{n}\sum_{i=1}^n J_i(\hat{\theta}) + o_p(n^{-1/2}) \quad \text{and} \quad \sum_{i=1}^n\tilde{\pi}_i(\hat{\theta})\psi_i(\hat{\theta})\psi_i'(\hat{\theta}) = \frac{1}{n}\sum_{i=1}^n\psi_i(\hat{\theta})\psi_i'(\hat{\theta}) + o_p(n^{-1/2}).$$

Because $\hat{\theta}$ is $\sqrt{n}$-convergent and by the regularity conditions in Assumption C.2 we have

$$\sum_{i=1}^{n} \tilde{\pi}_i(\hat{\theta}) J_i(\hat{\theta}) = E\left(J_i(\theta_*)\right) + O_p(n^{-1/2}) \quad \text{and} \quad \sum_{i=1}^{n} \tilde{\pi}_i(\hat{\theta}) \psi_i(\hat{\theta}) \psi_i'(\hat{\theta}) = E\left(\psi_i(\theta_*)\psi_i'(\theta_*)\right) + O_p(n^{-1/2}).$$

Then,

$$Z_n(\hat{\theta}) \equiv \tilde{G}(\hat{\theta})[\tilde{M}(\hat{\theta})]^{-1} \xrightarrow{p} Z(\theta_*) \equiv E\left(J_i'(\theta_*)\right)[\Omega(\theta_*)]^{-1}.$$

Next, we show that $\hat{\theta}^{m3s} \xrightarrow{p} \theta_{**}$ using Lemma A.1. We need to show that $\sup_{\theta \in \Theta} \|h_n(\theta) - h(\theta)\| \xrightarrow{p} 0$ with $h_n(\theta) = Z_n(\hat{\theta})\bar{\psi}(\theta)$ and $h(\theta) = Z(\theta_*)E\left(\psi_i(\theta)\right)$. Obviously,

$$\|h_n(\theta) - h(\theta)\| \leq \|Z_n(\hat{\theta})\| \|\bar{\psi}(\theta) - E\left(\psi_i(\theta)\right)\| + \|Z_n(\hat{\theta}) - Z(\theta_*)\| \|E\psi_i(\theta)\|.$$

By the same arguments as in the proof of Theorem 3.1 we can deduce that $\sup_{\theta \in \Theta} \|h_n(\theta) - h(\theta)\| \xrightarrow{p} 0$ and therefore, $\hat{\theta}^{m3s} \xrightarrow{p} \theta_{**}$ $\square$

Proof of Theorem 3.3. By the usual mean-value expansion,

$$\bar{G}_\pi(\hat{\theta}) = \bar{G}_\pi(\theta_*) + R_{p,q}\left(\frac{\partial vecG_\pi}{\partial \theta'}(\theta_*)(\hat{\theta} - \theta_*)\right) + O_p(n^{-1})$$

and

$$\bar{M}_\pi(\hat{\theta}) = \bar{M}_\pi(\theta_*) + R_{q,q}\left(\frac{\partial vecM_\pi}{\partial \theta'}(\theta_*)(\hat{\theta} - \theta_*)\right) + O_p(n^{-1})$$

$$\begin{aligned} \bar{G}(\hat{\theta})\bar{M}^{-1}(\hat{\theta})\bar{\psi}(\theta_{**}) &= \bar{G}_\pi(\hat{\theta})\bar{M}_\pi^{-1}(\hat{\theta})\bar{\psi}(\theta_{**}) \\ &= \left(\bar{G}_\pi(\hat{\theta}) - \bar{G}_\pi(\theta_*)\right) m_*\mu_{**} + \left(\bar{G}_\pi(\theta_*) - G_\pi(\theta_*)\right) m_*\mu_{**} \\ &\quad + G_\pi(\theta_*)\bar{M}_\pi^{-1}(\hat{\theta})\bar{\psi}(\theta_{**}) + O_p(n^{-1}). \end{aligned}$$

Since $a^{-1} - b^{-1} = -b^{-1}(a-b)a^{-1}$,

$$\begin{aligned} \bar{G}(\hat{\theta})\bar{M}^{-1}(\hat{\theta})\bar{\psi}(\theta_{**}) &= \left(\bar{G}_\pi(\hat{\theta}) - \bar{G}_\pi(\theta_*)\right) m_*\mu_{**} + \left(\bar{G}_\pi(\theta_*) - g_*\right) m_*\mu_{**} \\ &\quad - g_* m_* \left(\bar{M}_\pi(\hat{\theta}) - \bar{M}_\pi(\theta_*)\right) m_*\mu_{**} - g_* m_* \left(\bar{M}_\pi(\theta_*) - M_\pi(\theta_*)\right) m_*\mu_{**} \\ &\quad + g_* m_* \left(\bar{\psi}(\theta_{**}) - \mu_{**}\right) + O_p(n^{-1}). \end{aligned} \tag{C5}$$

The leading term in the expansions of $\bar{G}_\pi(\hat{\theta}) - \bar{G}_\pi(\theta_*)$ and $\bar{M}_\pi(\hat{\theta}) - \bar{M}_\pi(\theta_*)$ are

$$R_{p,q}\left(\frac{\partial vecG_\pi}{\partial \theta'}(\theta_*)(\hat{\theta} - \theta_*)\right), \quad \text{and} \quad R_{q,q}\left(\frac{\partial vecM_\pi}{\partial \theta'}(\theta_*)(\hat{\theta} - \theta_*)\right)$$

which are, up to $O_p(n^{-1})$, thanks to (15) and (16) linear functions of

$$\bar{J}(\theta_*) - J_*,\ \bar{\psi}(\theta_*) - \mu_*,\ \bar{J}(\theta_*^1) - J^*,\ \bar{\psi}(\theta^*) - \mu^*, \text{ and } \Omega_n(\theta_*^1) - \Omega(\theta_*^1).$$

Next, we expand $\bar{G}_\pi(\theta_*) - g_*$ and $\bar{M}_\pi(\theta_*) - M_\pi(\theta_*)$. It is more appropriate to proceed component-wise. Let X_{kl} be the (k,l)-component of the matrix X.

The (k,l)-component of $\bar{G}_\pi(\theta_*) - g_*$ is $\sum_{i=1}^n \pi_i(\theta_*)J_{i,lk}(\theta_*) - g_{*,kl}$ and we have

$$\begin{aligned}
\textstyle\sum_{i=1}^n \pi_i(\theta_*)J_{i,lk}(\theta_*) - g_{*,kl} &= -g_{*,kl} + \tfrac{1}{n}\textstyle\sum_{i=1}^n J_{i,lk}(\theta_*) - \bar{\psi}'(\theta_*)V_n^{-1}(\theta_*)\tfrac{1}{n}\sum_{i=1}^n \psi_i(\theta_*)J_{i,lk}(\theta_*) \\
&= -g_{*,kl} + \tfrac{1}{n}\textstyle\sum_{i=1}^n J_{i,lk}(\theta_*) - (\bar{\psi}(\theta_*) - \mu_*)'v_* E\left(\psi_i(\theta_*)J_{i,lk}(\theta_*)\right) \\
&\quad -\mu_*' V_n^{-1}(\theta_*)\tfrac{1}{n}\textstyle\sum_{i=1}^n \psi_i(\theta_*)J_{i,lk}(\theta_*) + O_p(n^{-1}) \\
&= -g_{*,kl} + \tfrac{1}{n}\textstyle\sum_{i=1}^n J_{i,lk}(\theta_*) - (\bar{\psi}(\theta_*) - \mu_*)'v_* E\left(\psi_i(\theta_*)J_{i,lk}(\theta_*)\right) \\
&\quad +\mu_*' v_* \left(V_n(\theta_*) - V(\theta_*)\right) v_* E\left(\psi_i(\theta_*)J_{i,lk}(\theta_*)\right) \\
&\quad -\mu_*' v_* \tfrac{1}{n}\textstyle\sum_{i=1}^n \left(\psi_i(\theta_*)J_{i,lk}(\theta_*) - E\left(\psi_i(\theta_*)J_{i,lk}(\theta_*)\right)\right) \\
&\quad -\mu_*' v_* E\left(\psi_i(\theta_*)J_{i,lk}(\theta_*)\right) + O_p(n^{-1}).
\end{aligned}$$

But

$$\begin{aligned}
V_n(\theta_*) - V(\theta_*) &= \tfrac{1}{n}\textstyle\sum_{i=1}^n \psi_i(\theta_*)(\psi_i(\theta_*) - \bar{\psi}(\theta_*))' - V(\theta_*) \\
&= \Omega_n(\theta_*) - \Omega(\theta_*) - (\bar{\psi}(\theta_*) - \mu_*)\mu_*' - \mu_*(\bar{\psi}(\theta_*) - \mu_*)' + O_p(n^{-1}).
\end{aligned}$$

Then

$$\begin{aligned}
\textstyle\sum_{i=1}^n \pi_i(\theta_*)J_{i,lk}(\theta_*) - g_{*,kl} &= \tfrac{1}{n}\textstyle\sum_{i=1}^n \left(J_{i,lk}(\theta_*) - J_{lk}(\theta_*)\right) - (\bar{\psi}(\theta_*) - \mu_*)'v_* E\left(\psi_i(\theta_*)J_{i,lk}(\theta_*)\right) \\
&\quad +\mu_*' v_* \left(\Omega_n(\theta_*) - \Omega(\theta_*)\right) v_* E\left(\psi_i(\theta_*)J_{i,lk}(\theta_*)\right) \\
&\quad -\mu_*' v_* (\bar{\psi}(\theta_*) - \mu_*)\mu_*' v_* E\left(\psi_i(\theta_*)J_{i,lk}(\theta_*)\right) \\
&\quad -\mu_*' v_* \mu_* (\bar{\psi}(\theta_*) - \mu_*)' v_* E\left(\psi_i(\theta_*)J_{i,lk}(\theta_*)\right) \\
&\quad -\mu_*' v_* \tfrac{1}{n}\textstyle\sum_{i=1}^n \left(\psi_i(\theta_*)J_{i,lk}(\theta_*) - E\left(\psi_i(\theta_*)J_{i,lk}(\theta_*)\right)\right) + O_p(n^{-1}).
\end{aligned} \tag{C6}$$

Thus, $\bar{G}_\pi(\theta_*) - g_*$ is a linear function of

$$\bar{J}(\theta_*) - J_*,\ \bar{\psi}(\theta_*) - \mu_*,\ \Omega_n(\theta_*) - \Omega(\theta_*),\ \text{and}\ \frac{1}{n}\sum_{i=1}^n \left(\psi_i(\theta_*) \otimes vec(J_i(\theta_*)) - E(\psi_i(\theta_*) \otimes vec(J_i(\theta_*)))\right).$$

Besides, the (k,l)-component of $\bar{M}_\pi(\theta_*) - M_\pi(\theta_*)$ is $\sum_{i=1}^n \pi_i(\theta_*)\psi_{i,k}(\theta_*)\psi_{i,l}(\theta_*) - (M_\pi(\theta_*))_{kl}$ and

$$\begin{aligned}
(\bar{M}_\pi(\theta_*))_{kl} - (M_\pi(\theta_*))_{kl} &= \Omega_{n,kl}(\theta_*) - \Omega_{kl}(\theta_*) \\
&\quad -(\bar{\psi}(\theta_*) - \mu_*)'v_* E\left(\psi_i(\theta_*)\psi_{i,k}(\theta_*)\psi_{i,l}(\theta_*)\right) \\
&\quad +\mu_*' v_* \left(\Omega_n(\theta_*) - \Omega(\theta_*)\right) v_* E\left(\psi_i(\theta_*)\psi_{i,k}(\theta_*)\psi_{i,l}(\theta_*)\right) \\
&\quad -\mu_*' v_* (\bar{\psi}(\theta_*) - \mu_*)\mu_*' v_* E\left(\psi_i(\theta_*)\psi_{i,k}(\theta_*)\psi_{i,l}(\theta_*)\right) \\
&\quad -\mu_*' v_* \mu_* (\bar{\psi}(\theta_*) - \mu_*)' v_* E\left(\psi_i(\theta_*)\psi_{i,k}(\theta_*)\psi_{i,l}(\theta_*)\right) \\
&\quad -\mu_*' v_* \tfrac{1}{n}\textstyle\sum_{i=1}^n \left(\psi_i(\theta_*)\psi_{i,k}(\theta_*)\psi_{i,l}(\theta_*) - E\left(\psi_i(\theta_*)\psi_{i,k}(\theta_*)\psi_{i,l}(\theta_*)\right)\right) \\
&\quad +O_p(n^{-1}).
\end{aligned} \tag{C7}$$

Thus, $\bar{M}_\pi(\theta_*) - M_\pi(\theta_*)$ is a linear function of

$$\Omega_n(\theta_*) - \Omega(\theta_*),\ \bar{\psi}(\theta_*) - \mu_*,\ \text{and}\ \frac{1}{n}\sum_{i=1}^n \left(\psi_i(\theta_*) \otimes vec(\psi_i(\theta_*)\psi_i'(\theta_*)) - E(\psi_i(\theta_*) \otimes vec(\psi_i(\theta_*)\psi_i'(\theta_*)))\right).$$

Therefore, from Equations (15), (16), (C5) (C6) and (C7), $\bar{G}(\hat{\theta})\bar{M}^{-1}(\hat{\theta})\bar{\psi}(\theta_{**})$ is, up to $O_p(n^{-1})$, a linear of

$$vec\left(\bar{J}(\theta_*) - J_*\right),\ \bar{\psi}(\theta_*) - \mu_*,\ \bar{\psi}(\theta_{**}) - \mu_{**},\ vec\left(\bar{J}(\theta_*^1) - J^*\right),\ \bar{\psi}(\theta^*) - \mu^*,\ vec\left(\Omega_n(\theta_*^1) - \Omega(\theta_*^1)\right),$$

$$vec\left(\Omega_n(\theta_*) - \Omega(\theta_*)\right),\ \frac{1}{n}\sum_{i=1}^{n}\left(\psi_i(\theta_*) \otimes vec(J_i(\theta_*)) - E(\psi_i(\theta_*) \otimes vec(J_i(\theta_*)))\right),$$

and

$$\frac{1}{n}\sum_{i=1}^{n}\left(\psi_i(\theta_*) \otimes vec(\psi_i(\theta_*)\psi_i'(\theta_*)) - E(\psi_i(\theta_*) \otimes vec(\psi_i(\theta_*)\psi_i'(\theta_*)))\right).$$

Let $\bar{\zeta} - \zeta_0$ be the vector obtained by stacking all of these centered sample means. Clearly,

$$\sqrt{n}(\bar{\zeta} - \zeta_0) \xrightarrow{d} \mathcal{N}(0, \Sigma)$$

and because of the linearity between the leading term in the expansion of $\bar{G}(\hat{\theta})\bar{M}^{-1}(\hat{\theta})\bar{\psi}(\theta_{**})$ and $\bar{\zeta} - \zeta_0$, there exists a matrix A of suitable size such that

$$\sqrt{n}\bar{G}(\hat{\theta})\bar{M}^{-1}(\hat{\theta})\bar{\psi}(\theta_{**}) = A\sqrt{n}(\bar{\zeta} - \zeta_0) + O_p(n^{-1/2}).$$

Therefore,

$$\sqrt{n}\bar{G}(\hat{\theta})\bar{M}^{-1}(\hat{\theta})\bar{\psi}(\theta_{**}) \xrightarrow{d} \mathcal{N}\left(0, A\Sigma A'\right)$$

and

$$\sqrt{n}\left(\hat{\theta}^{3s} - \theta_{**}\right) \xrightarrow{d} \mathcal{N}\left(0, D_*^{-1}A\Sigma A' D_*^{-1'}\right).$$

If the moment condition model is well specified, $\theta_0 \equiv \theta_* = \theta^* = \theta_{**}$, $\mu_* = \mu^* = \mu_{**} = 0$, $M_\pi(\theta_*) = \Omega(\theta_0)$ and $M_\pi(\theta_*) = J(\theta_0)$. From (C5),

$$\bar{G}(\hat{\theta})\bar{M}^{-1}(\hat{\theta})\bar{\psi}(\theta_{**}) = g_* m_* \bar{\psi}(\theta_{**}) + O_p(n^{-1})$$

and

$$\sqrt{n}\left(\hat{\theta}^{3s} - \theta_{**}\right) \xrightarrow{d} \mathcal{N}\left(0, (J_*'\Omega(\theta_*)^{-1}J_*)^{-1}\right)$$

which is the usual asymptotic distribution□

Proof of Theorem 3.4. Similarly to the expansion in (C5), we have

$$\begin{aligned}\bar{J}'(\hat{\theta})\Omega_n^{-1}(\hat{\theta})\bar{\psi}(\theta_{**}) \quad = \quad & \left(\bar{J}'(\hat{\theta}) - \bar{J}'(\theta_*)\right)\omega_*\mu_{**} + \left(\bar{J}'(\theta_*) - J_*'\right)\omega_*\mu_{**} - J_*'\omega_*(\Omega_n(\hat{\theta}) - \Omega_n(\theta_*))\omega_*\mu_{**} \\ & - J_*'\omega_*(\Omega_n(\theta_*) - \Omega(\theta_*))\omega_*\mu_{**} + J_*'\omega_*(\bar{\psi}(\theta_{**}) - \mu_{**}) + O_p(n^{-1}).\end{aligned} \tag{C8}$$

But

$$\bar{J}(\hat{\theta}) = \bar{J}(\theta_*) + R_{q,p}\left(J^{(2)}(\theta_*)(\hat{\theta} - \theta_*)\right) + O_p(n^{-1}),$$

$$\Omega_n(\hat{\theta}) = \Omega_n(\theta_*) + R_{q,q}\left(\frac{\partial vec[\Omega]}{\partial \theta'}(\theta_*)(\hat{\theta} - \theta_*)\right) + O_p(n^{-1}).$$

Hence from Equations (16) and (15), the leading terms of the expansions of $\bar{J}(\hat{\theta}) - \bar{J}(\theta_*)$ and $\Omega_n(\hat{\theta}) - \Omega_n(\theta_*)$ are linear functions of

$$\bar{J}(\theta_*) - J_*,\ \bar{\psi}(\theta_*) - \mu_*,\ \bar{J}(\theta_*^1) - J^*,\ \bar{\psi}(\theta^*) - \mu^*, \text{ and } \Omega_n(\theta_*^1) - \Omega(\theta_*^1).$$

Therefore, from (C8) the leading term of the expansion of $\bar{J}'(\hat{\theta})\Omega_n^{-1}(\hat{\theta})\bar{\psi}(\theta_{**})$ is a linear function of

$$vec(\bar{J}(\theta_*) - J_*),\ \bar{\psi}(\theta_*) - \mu_*,\ \bar{\psi}(\theta_{**}) - \mu_{**},\ vec(\bar{J}(\theta_*^1) - J^*),\ \bar{\psi}(\theta^*) - \mu^*,$$

$$vec(\Omega_n(\theta_*^1) - \Omega(\theta_*^1)), \text{ and } vec(\Omega_n(\theta_*) - \Omega(\theta_*)).$$

Let $\bar{\zeta}^m - \zeta_0^m$ be the vector obtained by stacking all of these centered sample means. Clearly,

$$\sqrt{n}(\bar{\zeta}^m - \zeta_0^m) \xrightarrow{d} \mathcal{N}(0, \Sigma^m)$$

and because of the linearity between the leading term in the expansion of $bar J'(\hat{\theta})\Omega_n^{-1}(\hat{\theta})\bar{\psi}(\theta_{**})$ and $\bar{\zeta}^m - \zeta_0^m$, there exists a matrix A^m of suitable size such that

$$\sqrt{n}\bar{J}'(\hat{\theta})\Omega_n^{-1}(\hat{\theta})\bar{\psi}(\theta_{**}) = A^m\sqrt{n}(\bar{\zeta}^m - \zeta_0^m) + O_p(n^{-1/2}).$$

Therefore,

$$\bar{J}'(\hat{\theta})\Omega_n^{-1}(\hat{\theta})\bar{\psi}(\theta_{**}) \xrightarrow{d} \mathcal{N}\left(0, A^m\Sigma A^{m'}\right)$$

and

$$\sqrt{n}\left(\hat{\theta}^{m3s} - \theta_{**}\right) \xrightarrow{d} \mathcal{N}\left(0, D_*^{m^{-1}}A^m\Sigma^m A^{m'}D_*^{m^{-1'}}\right).$$

Similarly to the proof of Theorem 3.3, it is easy to see that if the moment condition model is well specified,

$$\sqrt{n}\left(\hat{\theta}^{m3s} - \theta_{**}\right) \xrightarrow{d} \mathcal{N}\left(0, (J_*'\Omega(\theta_*)^{-1}J_*)^{-1}\right)$$

which is the usual asymptotic distribution$\square$

References

[1] Altonji, J. and Segal, L. (1996). Small Sample Bias in Gmm estimation of Covariance Structures. *Journal of Business & Economic Statistics* **14** 353–366.

[2] Andersen, T. G. and Sørensen, B. E. (1996). Gmm Estimation of a Stochastic Volatility Model: A Monte Carlo Study. *Journal of Business & Economic Statistics* **14** 328–352.

[3] Antoine, B., Bonnal, H. and Renault, E. (2007). On the Efficient Use of the Informational Content of Estimating Equations: Implied Probabilities and Euclidean Empirical Likelihood. *Journal of Econometrics* **138** 461-487.

[4] Back, K. and Brown, D. P. (1993). Implied Probabilities in Gmm Estimators. *Econometrica* **61** 971-975.

[5] Baggerly, K. A. (1998). Empirical Likelihood as Goodness-of-fit Measure. *Biometrika* **85** 535-547.

[6] Brown, B. W. and Newey W. K. (2002). Generalized Method of Moments, Efficient Bootstrapping, and Improved Inference. *Journal of Business & Economic Statistics* **20** 507-517.

[7] Chamberlain, G. (1987). Asymptotic Efficiency in Estimation with Conditional Moment Restrictions. *Journal of Econometrics* **34** 305-334.

[8] Corcoran, S. A. (1998). Bartlett Adjustment of Empirical Discrepancy Statistics. *Biometrika* **85** 967 972.

[9] Davidson, J. (1994). *Stochastic Limit Theory.* Oxford University Press, Oxford.

[10] Donald, G. S., Imbens G. W. and Newey W. K. (2002). Empirical Likelihood Estimation and Consistent Tests with Conditional Moment Restrictions. *Journal of Econometrics* **117** 55-93.

[11] Dovonon, P. and Renault E. (2008). Gmm Overidentification Test with First Order Underidentification. *Working Paper.*

[12] Guggenberger, P. (2008). Finite Sample Evidence Suggesting a Heavy Tail Problem of the Generalized Empirical Likelihood Estimator. *Econometric Reviews* **27:4** 526-541.

[13] Hall, A. R. (2000). Covariance Matrix Estimation and the Power of the Overidentifying Restrictions Test. *Econometrica* **68** 1517-1527.

[14] Hall, A. R. and Inoue A. (2003). The Large Sample Behaviour of the Generalized Method of Moments Estimator in Misspecified Models. *Journal of Econometrics* **114** 361-394.

[15] Hall, P. and Horowitz J. (1996). Bootstrap Critical Values for Tests Based on Generalized Method of Moment Estimators. *Econometrica* **64** 891-916.

[16] Hansen, L. P., 1982. "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029-1054.

[17] Hansen, L. P., Heaton P. and Yaron A. (1996). Finite-Sample Properties of Some Alternative Gmm Estimators. *Journal of Business & Economic Statistics* **14** 362-280.

[18] Imbens, G. W. (1997). One-Step Estimators in Overidentified Generalized Method of Moments Estimator. *Review of Economic Studies* **64** 359-383.

[19] Imbens, W. G. and Spady R. H. (2002). Confidence Intervals in Generalized Method of Moments Models. *Journal of Econometrics* **107** 87-98.

[20] Imbens, W. G., Spady R. H. and Johnson P. (1998). Information Theoretic Approaches to Inference in Moment Condition Models. *Econometrica* **66** 333-357.

[21] Kan, R. and Robotti C. (2008). Model Comparison Using the Hansen-Jagannathan Distance. *Review of Financial Studies*, Forthcoming.

[22] Kitamura, Y. (1997). Empirical Likelihood Methods with Weakly Dependent Processes. *Annals of Statistics* **25** 2084-2102.

[23] Kitamura, Y. (2001). Asymptotic Optimality of Empirical Likelihood for Testing Moment Restrictions. *Econometrica* **69** 1661-1672.

[24] Kitamura, Y. (2006). Empirical Likelihood Methods in Econometrics: Theory and Practice. *Cowles Foundation for Research in Economics, Yale University*, Discussion paper No. 1569.

[25] Kitamura, Y. and Stutzer M. (1997). An Information-Theorethic Alternative to Generalized Method of Moments Estimation. *Econometrica* **65** 861-874.

[26] Kitamura, Y., Tripathi G. and Ahn H. (2004). Empirical Likelihood-based Inference in Conditional Moment Restriction Models. *Econometrica* **72** 1667-1714.

[27] Magdalinos, A. M. and Symeonides S. D. (1996). A Reinterpretation of the Tests of Overidentifying Restrictions. *Journal of Econometrics* **73** 325-353.

[28] Mykland, P. A. (1995). Dual Likelihood. *Annals of Statistics* **23** 396-421.

[29] Newey, K. W. (1985). Generalized Method of Moments Specification Testing. *Journal of Econometrics* **29** 229-256.

[30] Newey, K. W. and McFadden D. (1994). Large Sample Estimation and Hypothesis. *Handbook of Econometrics*, **IV**, Edited by R. F. Engle and D. L. McFadden, 2112-2245.

[31] Newey, K. W. and Smith R. J. (2004). Higher Order Properties of Gmm and Generalized Empirical Likelihood Estimators. *Econometrica* **72** 219-255.

[32] Newey, K. W. and West K. D. (1987). A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica* **55** 703-708.

[33] Owen, A. (1990). Empirical Likelihood Ratio Confident Regions. *Annals of Statistics* **18** 90-120.

[34] Owen, A. (2001). *Empirical Likelihood*. Chapman & Hall, New York.

[35] Qin, J. and Lawless J. (1994). Empirical Likelihood and General Estimating Equations. *Annals of Statistics* **22** 300-325.

[36] Ramalho, J. S. and Smith R. J. (2002). Generalized Empirical Likelihood Non-Nested Tests. *Journal of Econometrics* **107** 99-125.

[37] Robinson, P. M. (1988). The Stochastic Difference Between Econometric Statistics. *Econometrica* **56** 531-548.

[38] Schennach, S. M. (2007). Point Estimation with Exponentially Tilted Empirical Likelihood. *Annals of Statistics* **35** 634-672.

[39] White, H. 1982. Maximum Likelihood Estimation of Misspecified Models. *Econometrica* **50** 1-25.